

'Gene-foods': the alternative to a moratorium

Three developments could help reduce public concern about genetically modified crops: broader criteria to assess their social and environmental impact, an over-arching monitoring body, and greater transparency.

A wave of protest in recent weeks against genetically modified crops in Britain seems to have taken scientists, industrial managers and politicians alike by surprise. After all, Britain already has one of the most highly developed regulatory systems in Europe to assess the potential threat to the environment of experimental tests of such crops. This includes formal mechanisms both for obtaining scientific advice about the likelihood and potential size of such threats, and for ensuring consultation with local communities before the tests are carried out. Despite this, newspaper reports last week suggested that almost 150 groups have formed throughout the country pledged to 'non-violent' action against such experiments. What has gone wrong? And what can be done to remedy the situation?

It is tempting — but ultimately misleading — to put the problem down to a lack of understanding, either of the scientific processes of gene transfer, or of its potential value to agriculture. The first explanation appeals to the widely discredited 'deficit model' of the public understanding of science. This suggests that distrust of the products of modern technology is based primarily on a lack of sufficient information on its scientific basis. In practice, additional information is as likely to fan distrust — for example, by helping to focus on the scientific uncertainties involved in making calculations of risk — as to dispel it.

The second explanation ignores the extent to which concern about modern agricultural production — as the rising popularity of organic food demonstrates — is as much about the quality of produce as about its quantity or price.

Protests

Admittedly, there has been a strong unscientific current in much of the recent debate. The protesters who last week destroyed a trial of genetically modified forage maize at a farm near Dartington in Devon (see page 608) justified their actions by the claim that the crop could pollinate a crop of organic sweetcorn almost two kilometres away. But the chances of contamination at this distance had already been assessed, and dismissed, by the government's Advisory Committee on Releases to the Environment, in a decision upheld by the Court of Appeal.

The past few years have seen a growing wealth of scientific knowledge — much of it reassuring — about the likelihood of implanted genes being transferred between novel crops and their surrounding environment. This in itself is sufficient reason for rejecting the widely heard demand for a blanket moratorium on the commercial planting of genetically modified crops; such moves can only be justified by novel and unexpected scientific developments — such as last year's news on the possibility of cloning adult humans — or by clear evidence of a potential danger. Banning the cultivation of certain crops at the same time as permitting their import would be hypocritical.

Conversely, however, proponents of genetically modified crops need to acknowledge the existence of significant scientific uncertainties about their potential ecological impact. Last week, for example, studies described at the Ecological Society of America were reported to have shown that weeds growing close to herbicide-tolerant crops,

which had themselves developed herbicide resistance as a result, remained as healthy as in their uncontaminated state. Similarly, earlier this week, the Rowett Research Institute in Aberdeen announced preliminary findings suggesting that genetically modified, insect-resistant potatoes can damage the immune system in rats. It is clear that much remains to be understood about the precise interaction between genetically modified crops and their immediate environment.

Concerns

It is also clear that the recent protests, far from being based on a simple or deliberate lack of understanding, have a more complex relationship to a broad set of concerns about modern food production, including the degree of trust placed in regulatory authorities.

One possible explanation for the relative lack of public opposition to genetically engineered crops in the United States is the high level of public confidence in which the Food and Drug Administration is held. In contrast, the failures of Britain's regulatory authorities over bovine spongiform encephalopathy — widely blamed on the fact that regulatory responsibility was in the hands of a government ministry whose first priority is to promote agricultural production — appear to be coming home to roost.

Three specific proposals can be made to meet this distrust. The first is to ensure that, when a proposed new crop is assessed for its potential negative impacts, the scope of any such assessment is cast sufficiently broadly to ensure that an adequate range of public concern is taken into account. Assessments limited to narrow scientific or technical questions, essential as these are, will not necessarily answer broader questions, such as the implications of changing farming practices for local wildlife or for eating habits.

The second proposal is that a new, over-arching body, independent of both government and the food industry, be set up to monitor the development of genetically modified foodstuffs and their social and environmental impacts. Adding a new layer of bureaucracy is always a step to be taken with caution. But such a body, to which the current plethora of scientific advisory committees with an interest in the field would report regularly, and which in turn would ensure that key strategic objectives do not become lost in a fragmentation of regulatory responsibilities, could perform an invaluable exercise in restoring public trust. The Human Genetics Advisory Commission already provides a successful and appropriate model.

Finally, it is important that, if trust in the regulatory process is to be built and maintained, those responsible for the current wave of protests must feel that their points of view are not only heard but also listened to. This does not mean that they have to be accepted, or the protesters' current tactics endorsed. But where consumer representatives are placed either on scientific advisory boards, or on a higher monitoring body, these should not be token appointments, but able to represent the expressed point of view of consumers. At the same time, the deliberations of such panels must be transparent if public confidence in their integrity and independence is to be assured. □



Radioastronomers hammer out an agreement on mobile phones

[MUNICH] Radioastronomers have gained a small victory in their battle with the telecommunications satellite operator Iridium over use of an important radiofrequency. But victory comes at a cost: they have had to accept a significant loss in observing time.

Iridium launches its global mobile telephone service next month. It is based on a system of 66 satellites operating both uplinks and downlinks in a single frequency band, adjacent to the 1,612 MHz band reserved by international law for radioastronomers.

Technical inadequacies in Iridium's system mean that overspill emissions from the downlink will drown out the weak 1,612-MHz signals from deep space (*Nature* 380, 569; 1996).

While US radioastronomers have reached a time-sharing agreement with Iridium, their European counterparts have fought hard to protect their frequencies. They have won the support of their governments, which have withheld operating licences from Iridium until agreement is reached over acceptable pollution levels.

After six months of acrimonious negotiations, Iridium this week signed an agreement with the European Science Foundation (ESF) on behalf of its associated Committee on Radio Astronomy Frequencies (CRAF). This should keep the 1,612 MHz band clear of interference until next March. The situation should improve after January 2006, the deadline by which Iridium must replace its 66 satellites with an improved non-polluting system.

According to the terms of the agreement, Iridium and radioastronomers will have to work together to develop non-polluting next-generation satellites, and to develop technical fixes to reduce the susceptibility of radioastronomy equipment to overspill emissions.

During the next few months Iridium and radioastronomers will have to reach a further agreement on time-sharing during the period between March 1999 and 2006. With both sides determined to keep the lion's share of the time, the battle will be hard fought.

Willem Baan, chairman of the Inter-Union Commission on Frequency Allocations for Radioastronomy and Space Science, and director of the Westerbork Observatory in the Netherlands, is worried that the requirements of radioastronomers are more than Iridium will want to concede. He is also unhappy about the spirit of the negotiations so far. "Iridium knew of the radioastronomy problem as early as 1991," he says. "Yet it did little to solve it in the design phase of their system and so now we have a serious problem."



Close call: next March these dishes at Westerbork will begin competing with mobile phones.

He believes that radioastronomers are paying for Iridium's failure to solve its own problems. "It is still not clear that radioastronomers will be able to perform very sensitive observations in the immediate vicinity of the strong main transmission of Iridium," he says. "Other first-generation mobile satellite service systems avoid these problems by using other downlink frequencies."

Industrialist to head UK research councils

[LONDON] The British government has announced that John Taylor, director of Hewlett-Packard's European research centre in Bristol, will be the new Director General of Research Councils (DGRC). Taylor will take over on 1 January 1999 for a three-year term of office.

Taylor, who is also president-elect of the Institution of Electrical Engineers, has worked for Hewlett-Packard since 1984. His appointment continues a tradition set by the previous, Conservative government of choosing a scientist from industry rather than academia for the job of DGRC. The present incumbent, Sir John Cadogan, was formerly with the oil company BP.

Peter Mandelson, newly-appointed Secretary of State for Trade and Industry, and cabinet minister with responsibility for science, said he was "delighted" to see someone of John Taylor's stature as DGRC. "I passionately believe science is the bedrock to economic success," he said last week.

The move underlines the government's commitment to Foresight, an initiative to improve links between science and industry, set up by the previous government to harness science in the service of wealth creation and quality of life. Taylor is

Not only will radioastronomers be forced to accept time-sharing, but, even with a guarantee of 'no pollution' after 2006, they will lose about eight per cent of observation time as Iridium's armada of satellites pass directly across the piece of sky they are looking at. This is a loss that any telecommunications company would find unacceptable, says Baan.

However, ESF secretary general Enric Banda puts a positive spin on the agreement, which "underlines the willingness of Europe's radioastronomers to work constructively with the growing number of satellite companies, to find technical solutions that will allow science and industry to continue to profitably coexist in space".

But interference from satellites remains an increasing threat. "This is not an isolated problem," says Jim Cohen, CRAF's chairman. "Unless the protection of radioastronomy is taken into account early in the design of new satellite systems, our science could face a difficult future." Even with full cooperation from the telecommunications industry, the sheer number of satellites in systems like Iridium will eventually add up to a serious physical obstruction in the radio sky. **Alison Abbott**



Taylor: to theme or not to theme?

committed to Foresight as chair of the expert panel on information technology and communications.

One of Taylor's responsibilities will be to advise on the next phase of Foresight. Consultation is already under way on plans to put more emphasis on using science to enhance

quality of life. One suggestion is to reduce the numbers of expert panels and concentrate on getting science to focus on 'themes' such as ageing, crime control and the future of cities.

Taylor has criticized this approach (see *Nature* 393, 8; 1998), arguing that it is too early in the process to make such a change, particularly as many sectors of industry have yet to fully exploit the wealth-creating potential of better links with science.

Taylor's appointment completes a string of significant changes to the administration of British science, including the arrival of the industrialist Lord (David) Sainsbury in a newly-created junior ministerial post for science (see *Nature* 394, 511; 1998).

Ehsan Masood

Science returns to Einstein's rural retreat

[MUNICH] A six-year dispute over the ownership of Einstein's summer house in east Germany appears to have been resolved — to the benefit of science.

The Amt zur Regelung offener Vermögensfragen (ARV), the German office which establishes ownership of property confiscated by the Nazis, has ruled that the house should be given to Einstein's descendants.

This means it will almost certainly be preserved exclusively for scholarly activities, and no longer be used as a tourist attraction. It reverses the ARV's 1992 decision to give ownership to the tourist-hungry village of Caputh, where the house is located — though Caputh is planning to appeal, and objections from some beneficiaries about the rights of others to be considered heirs could still be lodged.

The wooden house, tucked in the forests of Brandenburg, has had a dismal history. Einstein spent only three summers there before leaving Germany permanently in 1933 when the Nazis rose to power. In 1935 Einstein was declared an enemy of the state and his properties were confiscated. The ownership of the summer house was transferred to the village of Caputh for DM5,000, a fraction of the price that Einstein had paid to build it. However, the village has not been able to prove that it actually paid the money.

Caputh initially used the house as a Jewish orphanage, but the children were driven out to die in the forests in 1938. Next, it was used as a camp by Nazi youth groups. After the war it was used to shelter refugees. It then fell into disrepair, but east German physicists persuaded the science ministry to renovate it in 1979 for Einstein's centenary. Thereafter it was used as a guest house for scholars by the DDR Academy of Sciences.

When the academy was dissolved after reunification, the house, now a protected historical monument, was entrusted to the state of Brandenburg until rightful ownership could be established. Hoping to continue the house's academic associations, Brandenburg set up the Einstein Forum, which it hoped could establish a series of scientific workshops and seminars there.

But the plan went awry when in 1992 the ARV declared Caputh the rightful owner, agreeing that the village had bought it legally in the 1930s. This decision was disputed by Einstein's heirs and the forum, who feared using it as a tourist attraction would endanger its fragile structure. The village made an agreement with Brandenburg and the Einstein Forum to restrict guided tours to weekends, in exchange for much-needed financial help. On top of standard running costs, there is an annual bill of DM140,000 for round-the-clock security services: the house was declared a 'politically endangered site' by the



Domestic science: tourism will make way for scholarly pursuits at Einstein's old summer house.

Brandenburg government after neo-nazis firebombed some properties in the area.

The forum is now trying to raise some DM2 million (US\$1.12 million) to buy the house from the 12 named beneficiaries, who are willing to sell, provided the house is used for scholarly purposes. One, the Hebrew University in Jerusalem, agreed to transfer its rights to the Einstein Forum. But Gary Smith,

president of the forum, is not prepared to guess when the house will find lasting peace. "Six years ago I predicted the issue would be resolved within half a year," he says, "so I won't be drawn on making predictions again."

Smith is keen to see the issue resolved quickly. The dispute has discouraged Caputh from investing in the house, which is falling once again into disrepair.

Alison Abbott

Crop trials speed up as eco-warriors strike

[LONDON] The British government, farmers, and the biotechnology industry are considering how to respond to destruction of genetically modified crop trials by 'eco-warrior' environmentalist groups.

Possible options include greater security around trial sites; carrying out trials in closed, company-owned locations, rather than fields owned by farmers; and simply not publicizing their specific location in a field.

Concern is mounting following an incident last week when protesters destroyed a crop of genetically modified maize in Devon, causing losses of £600,000 (US\$978,000). The crop was being assessed for a commercial seed listing, having passed safety tests.

The trial was at the centre of an unsuccessful court challenge by a coalition of groups including a nearby farmer and Friends of the Earth. They wanted it stopped, partly out of concern that it might pollinate a nearby crop of organic sweetcorn (see *Nature* 394, 212; 1998).

John McLeod, director of the National Institute for Agricultural Botany (NIAB), which was carrying out the Devon maize trial, is opposed to carrying out trials behind closed doors. He says they need to be carried out in open environments, particularly where pollination is being measured.

But while he remains committed to openness and transparency, McLeod says he is now beginning to have reservations about publicizing

the specific location of genetically modified trials. The law requires the government to publicize the location of a crop trial. However, there is no requirement to make the exact field grid reference public knowledge.

Government sources say they have considered changing their current practice of making specific crop sites public knowledge.

But the sources, as well as the company which was developing the Devon maize, Sharpes International Seeds, say this would be a retrograde step. "Having been so open now, we can't possibly go back," says a government official.

The Devon destruction is the latest in a string of incidents involving a new and relatively unknown group by the name of GenetiX Snowball, comprising environmental activists and local residents opposed to genetic modification in agriculture.

In the meantime, the Ministry of Agriculture, Fisheries and Food is to speed up the seed certification process by waiving the requirement on companies to show results of two of their own seed trials before submitting seed for a government-monitored trial.

A ministry spokesman said that this did not have implications for safety, which would already have been assessed by the Department of the Environment before any seed was passed onto the agriculture ministry for certification. **Ehsan Masood & Katherine Akingbade**

Dilemma for journals over tobacco cash

[WASHINGTON] Editors of medical journals have given mixed reactions to last week's revelations that more than a dozen scientists received \$156,000 from the tobacco industry to write letters and manuscripts disputing the carcinogenicity of second-hand smoke.

One of the targeted journals, the *Journal of the National Cancer Institute* (JNCI), is expected to change its policy and require future authors to divulge whether they have been paid by outside interests for a particular letter, and whether those interests reviewed and edited the letter.

Barnett Kramer, the journal's editor-in-chief, says the proposed change in policy aims "to be sure that we don't miss what a reasonable reader could interpret as a potential conflict".

But George Lundberg, editor of another targeted publication, the *Journal of the American Medical Association* (JAMA), says he is not troubled by the fact that a scientist was paid by the tobacco industry to write to his journal. This is because the journal acknowledged in a footnote that the Tobacco Institute provided support for the analysis contained in the letter. Lundberg says: "Our policy worked well, all sides were heard and the readers were informed."

The letters were written in 1992 and 1993, just before and soon after the Environmental Protection Agency (EPA) published a controversial report declaring second-hand smoke a dangerous carcinogen. They appeared in journals including JNCI, *The Lancet*, JAMA and *Pediatrics*.

The authors were paid \$2,250 to \$10,000 per letter by the Tobacco Institute, according to formerly confidential tobacco industry documents, which also reveal that two law

firms representing the industry revised some letters before submission. The documents were made public during a lawsuit brought by the state of Minnesota against the industry.

The revelations were first reported in the *St Paul Pioneer Press*, a Minnesota newspaper. The Tobacco Institute declined to comment. The law firms — Covington and Burling of Washington and Shook, Hardy & Bacon of Kansas City, Missouri — could not be reached for comment.

Four of five letters published in JNCI noted that their authors received support from the institute, or had done so in the past, as did a letter to JAMA for which Nathan Mantel, a biostatistician at the American University in Washington, DC, was paid \$10,000.

But there was no acknowledgement of support on a letter from Gio Batta Gori, a former deputy director of the Division of Cancer Causes and Prevention at the NCI, published in *The Lancet* on 10 April 1993, for which Gori was paid \$4,000. It attacks the January 1993 EPA report as being shoddy sci-

ence and concludes that "these are no trivial partisan questions in defence of tobacco interests". *The Lancet* has since changed its policy to require declaration of conflicts by authors.

Richard Horton, editor of *The Lancet*, says his reaction is one of "disgust". Gori, he says, "has breached a bond of trust as a scientist between himself and the scientific community". This, adds Horton, "is at best unethical and at worst an example of research misconduct".

Horton draws a distinction between an author receiving acknowledged support for continuing work and entering into a *quid pro quo* letter contract. "This is a project to seed the literature with a particular point of view for a defined sum of money... That's what I find so appalling."

The documents state that Gori, a consultant in Bethesda, Maryland, received \$20,137 from the Tobacco Institute for five letters. But Gori says the fact that he was paid for each letter is irrelevant: "The scientific issue is: are we right or are we wrong?"

Any attack is "libellous", he adds, "because it does not concern the substance of the matter but rather resorts to the time-honoured trick of killing the messenger when you don't like the message".

According to industry documents, the institute's strategy focused on recruiting scientists "at or near retirement with no dependence on grant-dispensing bureaucracies".

A federal judge in North Carolina ruled last month that the EPA report was based on inadequate science and was wrong in declaring environmental tobacco smoke a dangerous carcinogen. The EPA is to appeal against the decision.

Meredith Wadman



Researchers accused of wanting to rob the poor to give to NIH

[WASHINGTON] A leading organization representing US biomedical researchers has been harshly criticized by a senior Democrat, reflecting the increasingly polarized atmosphere surrounding a bill that aims to boost the budget of the National Institutes of Health (NIH) by 9.1 per cent next year.

Congressman David Obey (Democrat, Wisconsin), the senior Democrat on the House Appropriations Committee, is charging the Federation of American Societies for Experimental Biology (FASEB) with calling for a Republican-backed \$1.24 billion NIH increase at the expense of the poor. FASEB consists of 17 societies representing 56,000 researchers in the life sciences.

In a letter of 28 July to William Brinkley, president of FASEB, Obey writes that the group has "clearly crossed the line between advocating for your own interests and advo-

cating against the interests of others whose cause is just as imperative. That the 'others' in this instance are the weakest and most vulnerable members of society makes your position that much more disappointing."

Obey's letter was prompted by one from Brinkley to members of Congress urging them to "take whatever... actions are necessary" to ensure passage of the 1999 bill that would boost NIH funding to \$14.86 billion.

Democrats are strongly opposed to the bill because it contains \$2 billion in cuts to social programmes, including the elimination of a \$1.1 billion programme of home heating for the poor and a \$871 million summer job training programme for youths (see *Nature* 394, 108 & 311; 1998).

Obey appears to have read the FASEB letter as an endorsement of these cuts. "I don't believe your members want the Congress to

move in the direction your letter is advocating," Obey wrote to Brinkley. "[Biomedical researchers] are not men and women who wish to see low income elderly forced to choose between heating and eating."

But Brinkley has denied Obey's interpretation of his letter. In a response to Obey last week he says that his letter was not intended "to advocate against the interests and programs of other communities".

Brinkley says he was suggesting that Congress find more money to add to the bill, to restore the cuts to social programmes and to boost the NIH budget. "We don't want to be partisan in any way. We felt that we should support the bill for its broad-based support of health and biomedical research."

Republican leaders aim to bring the bill to the House floor within two weeks of reconvening on 8 September.

M. W.

Visa wrangle brings crisis in science jobs

[WASHINGTON] A number of non-US research staff hired by American universities are stranded without visas to enter the country following political stalemate over quotas for overseas-born professionals.

The visa problem is holding up temporary faculty appointments at a number of institutions, according to Sang Han, who tracks immigration issues for the Association of American Universities (AAU).

"At least three department heads are tearing their hair out" because non-US faculty members who were supposed to begin teaching classes in September will not arrive in time, says Richard Lariviere, vice-president of international programs at the University of Texas.

Visa applications in the 'H-1B' category — which includes computer programmers, physical therapists, and university jobs ranging from visiting professors to research associates — have been frozen since May, when the annual quota of 65,000 visas was filled.

A bill proposed by Congressman Lamar Smith (Republican, Texas) would add another 20,000 slots for this fiscal year, which ends on 30 September. But the bill was withdrawn last week, under threat of a presidential veto. The earliest Congress can take up the matter now is 9 September, when the House of Representatives returns from a month-long recess.

The controversy has its roots in the politics of the labour market for computer programmers and systems analysts, rather than academic scientists. The US software industry has lobbied strenuously for an increase in the number of H-1B visas, claiming that a shortage of native-born workers forces it to hire programmers from India and other countries.

But opponents of an H-1B increase — including some labour unions — say there is no domestic labour shortage. They say Silicon Valley companies are simply trying to hold down wages by hiring lower-paid staff from overseas.

The White House is torn between its desire to serve the politically powerful software industry, and its desire to side with workers' organizations. To become law, Smith's bill is likely to need stronger protection for US workers.

Universities are caught in the crossfire. They filed only two per cent of the 400,000 applications for H-1B visas last year (40 per cent of the applications to the Department of Labor were for programmers and 25 per cent for physical therapists).

Lariviere says the academic community is more discriminating than the software industry. "These [faculty members] are really very carefully targeted individuals," he says. "Our mandate is to find the best per-

son," unlike the software industry which, he claims, fills "job shops" with programmers holding temporary visas.

Smith's proposed legislation, a version of which has already passed the Senate, would raise the H-1B quota to 95,000 next year, 105,000 in 2000, and 115,000 in 2001 and 2002, before dropping to 65,000 again in 2003.

But, even if these increases are approved, Lariviere worries that they will only delay the inevitable, and that the H-1B slots will still run out before the end of the year. "I suspect all we've done [with the bill] is move back slightly the date when the cap will be hit."

Geneticists told to debate with politicians

[BEIJING] Geneticists throughout the world should engage more deeply in discussion with their governments on the implications of their work, and not stay aloof from such debates, one of the leading figures in the US Human Genome Project told an international meeting this week.

Maynard Olson, a professor of genetics at the University of Washington in Seattle, made his appeal in a plenary address to the Eighteenth International Congress of Genetics, which is being held in Beijing under the sponsorship of the Chinese Academy of Sciences and the Genetics Society of China.

The meeting is the third such congress to be held in Asia, after Tokyo in 1968 and New Delhi in 1983. The decision to hold it in China, which was made at the last congress, in Birmingham, United Kingdom, in 1993, has been controversial ever since China drafted a 'eugenics' law (see *Nature* 372, 123; 1994).

This law, which came into force in 1995, requires physicians to recommend couples to postpone marriage if one partner is found to have a series of contagious or mental diseases. Those diagnosed with a "genetic disease of a serious nature" are asked to take unspecified long-term contraceptive measures or to be sterilized.

After the decision to hold the meeting in Beijing had been confirmed, three national genetics associations — those in the United Kingdom, Argentina and the Netherlands — withdrew in protest from the International Genetics Federation (IGF), under whose auspices the regular international congresses are held, advocating a boycott by their members.

Other associations argued in favour of attending the meeting, and rejected the call for a boycott. But several did so only after they had been promised that the programme would include an open debate

In addition, universities fear that the software industry will quickly swallow up any new slots, steadily reducing the number of visas left for universities. Extra programmers are being recruited to help solve the 'millennium bug' problem.

Applications for H-1B visas are 'first-come, first-served,' with no quotas for any category of worker. The ceiling is reached earlier every year. In 1997 all 65,000 slots were taken by July, but this year they lasted only until May.

Han of the AAU says that, if nothing changes, next year's allocation of visas may all be gone by January.

Tony Reichhardt

with Chinese geneticists about the Maternal and Infant Health Care law and the potential dangers of its rigid application.

Although the local organizing committee in Beijing subsequently agreed to a request from the executive committee of the IGF to include a symposium on eugenics, this has been broadened into a general session on the ethical, legal and social implications of genetics research.

However, it will be followed by an evening workshop on the science and ethics of eugenics, chaired by Anthony Griffiths, professor of botany at the University of British Columbia, Canada, the secretary of the IGF, and Renzhong Qiu of the Chinese Academy of Social Science.

The eugenics issue was not referred to in the opening ceremonies of the congress, which concentrated on its main theme, "Genetics — Better life for All". The president of the congress, veteran Chinese geneticist C. C. Tan, emphasized how genetics can improve the quality of life, for example by helping to increase food production.

"With the world's population estimated to increase from 5.9 billion to more than 7 billion in 100 years' time, we face the question of whether the human species can sustain itself on Earth," said Tan. "Food production has to keep up with population growth, and international cooperation is clearly indispensable in developing new plant breeding skills."

Discussing the international implications of human genome analysis, Olson argued that it was wrong for geneticists or governments to spend too much energy on the relatively short-term issue of intellectual property rights over genetic material. "The issue that all countries should focus on is the development of their local scientific capacity," he said.

David Dickson

Fermilab faces up to uncertain future

Colin Macilwain

John Peoples will retire next year as director of Fermilab, leaving it in solid shape. But with CERN expected to usurp its position at the forefront of high-energy physics, and doubts about government investment, his replacement must find a way of ensuring a secure future for the US lab.

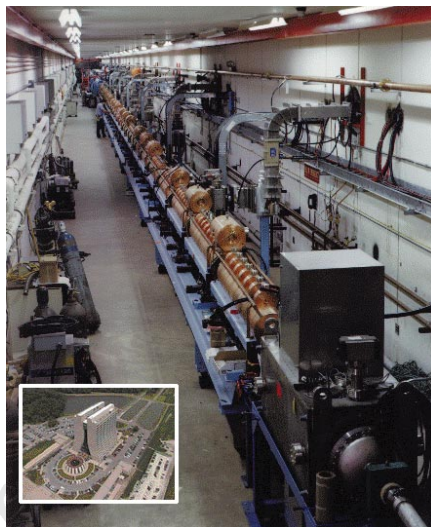
[BATAVIA, ILLINOIS] Wilson Hall, the modernist, 15-floor citadel that houses most of the physicists at the Fermi National Laboratory, brings an air of permanence to the laboratory that many more hastily constructed government facilities sorely lack. But in a few years' time that permanence will be tested.

A search is underway for a successor to John Peoples, who will retire as director of the largest US particle physics laboratory next summer. The challenge for the new director will be to find a role for Fermilab in six to eight years' time, when the Large Hadron Collider (LHC) at CERN, the European laboratory for particle physics, effectively usurps its position at the frontier of high-energy physics.

The director's appointment is crucial, says Peter Rosen, head of the high-energy physics programme at the Department of Energy, which funds the laboratory. "The person coming in will have to develop a clear vision for Fermilab beyond 2005," he says. "That's an important issue for high-energy physics — not just in the United States, but on a global scale, too."

In the short term, Fermilab's prospects are bright. Construction of a new main injector is to be completed this year, enabling its accelerator ring, the Tevatron, to resume operations next year with a far higher beam intensity than before. Two of the three main components of the \$530 million US contribution to the LHC are led from the Illinois laboratory. And several medium-sized projects are under way, including the Neutrinos at the Main Injector experiment, which would fire neutrinos at an underground target 500 miles north of the lab in a disused iron mine in Minnesota.

But, in the longer term, the challenge facing the laboratory is substantial. Since the 1993 abandonment of the Superconducting Super Collider in Texas, doubts persist about the willingness of the US government to pursue future multi-billion dollar investments in high-energy physics. In any case, the High Energy Physics Advisory Panel, following a report completed in February by a subpanel chaired by Fred Gilman, has called for a conceptual design report on a 1 TeV electron-positron linear collider, placing this Next Linear Collider (NLC) at the head of its



Straightforward future? Physicists from Wilson Hall (inset) operate the Tevatron, which receives protons boosted by the linear accelerator (above).

list of possible investments.

But as Ken Stanfield, Fermilab's deputy director, readily concedes, electrons are not Fermilab's area of expertise. The NLC will be built in Japan or at the Stanford Linear Accelerator Center in California, or not at all. Some Fermilab physicists predict that its costs will run too high, providing an opportunity for a rival proposal. One of the physicists, Bill Foster, suggests that the NLC costs could surpass \$3 billion and that an alternative that would advance particle physics at less cost could move to the front of the queue. "The first [team] that comes in with a \$2.5 billion machine" will be best placed to get there, Foster predicts.

Following the Gilman report, steps are being taken to refine two such approaches, either of which could be built at Fermilab. A steering committee has been established to plan for a future Very Large Hadron Collider (VLHC). The secretary of the committee, Fermilab physicist Ernest Malamud, says it will form three working groups — dealing with accelerator physics, accelerator technology, and magnets — to explore the technical challenges of designing a ring that would give proton energies of up to 100 TeV.

The working groups, open to scientists from all over the world, will explore both the

possible approaches to a VLHC — a huge ring with a low magnetic field, sometimes known as the Pipetron (see *Nature* 385, 471; 1997), and a compact ring with a very high magnetic field. The working groups will concentrate on cost issues, Malamud says.

The other option is for a muon collider. Recognizing the technical and fiscal constraints on electron and proton colliders, this approach would instead try to make use of the muon, a subatomic particle that decays in two microseconds if left alone, but whose properties would allow for very high-energy collisions in a ring of manageable size.

"We don't know if muon colliders are feasible or not," admits Steve Geer, a Fermilab physicist working on a research collaboration devoted to muon collider concepts involving the Illinois laboratory, the Brookhaven National Laboratory in New York and the Lawrence Berkeley National Laboratory in California. Muons and anti-muons would be generated by bombarding a target with protons. The team is tackling technical challenges, such as designing a target to survive the bombardment of protons, and finding a way to compress the resultant cloud of muons and anti-muons and accelerate them round a ring in opposite directions before they decay. If such a collider were feasible, it could reach very high energies in a small space, and two generations of machine could fit into Fermilab.

"The muon collider could have a lot more energy than the LHC and be accommodated on this site," says Stanfield. Asked if a major new accelerator will ever be built at Fermilab, Stanfield admits to some uncertainty, citing political problems that the lab will confront if it attempted to tunnel off-site.

Much of the lab's effort is already devoted to the US contribution to the LHC. Fermilab leads the US contribution to the Compact Muon Solenoid (CMS), one of two targets being designed for the LHC (the US contribution to ATLAS, the other target, is being coordinated at Brookhaven). The \$200 million US contribution to the construction of the accelerator itself is also managed at Fermilab, although much of the money will be spent with outside industrial contractors.

Each detector project has been reviewed by a team jointly appointed by the Department of Energy and the National Science Foundation, the other agency supporting the US contribution. In the case of the CMS, the outcome was that the proposed scope of the detector will be curtailed to allow the level of contingency funding — 40 per cent — which the reviewers believe is needed to make absolutely certain that the project remains within budget. Stanfield, who has overall management responsibility for the US contribution to the CMS, says the contingency money can be used later to restore the detector to its original specification, if it isn't needed to cover unexpected costs. □

FERMILAB

Expert panel will advise US officials on genetic testing

[WASHINGTON] Donna Shalala, the US Secretary of Health and Human Services, announced last week that she is creating an Advisory Committee on Genetic Testing to guide her on "broad, department-wide policy issues raised by genetic testing".

The panel of 11 independent experts should be appointed by the end of the year. It will be administered through the Office of Science Policy in the director's office at the National Institutes of Health.

"We need to ensure that test results are accurate and medically valid and that this information remains confidential," Shalala said in a statement. "This new advisory commission creates the mechanism that will help us make sure patients are protected."

Six non-voting members will represent the agencies of the Public Health Service, as well as the Health Care Financing Administration. David Satcher, the Assistant Secretary for Health, will play a key role in guiding the group's agenda.

Investors welcome new boss at British Biotech

[LONDON] British Biotech last week picked Christopher Hampson, a former director of Imperial Chemical Industries, to succeed John Raisman as non-executive chairman. The move was welcomed by investment analysts, one of whom described Hampson as "a solid, heavyweight industry figure".

Hampson acknowledged that he had had doubts about joining British Biotech. But "having met the people and understood their technology, I believe they have a technological lead. My challenge is to convert that into value for shareholders."

British Biotech shares fell following doubts about the efficacy of two drugs being developed by the company. Several senior staff have departed following a torrent of adverse publicity.

Modified maize escapes United Kingdom ban

[LONDON] The British government decided last week not to ban cultivation of a Novartis variety of genetically modified maize following advice from one of its scientific advisory committees.

Michael Meacher, the environment minister, had asked the committee on releases to the environment to assess a laboratory study showing that lacewing butterflies could be harmed when fed on prey that had eaten the maize. The maize has been modified to release an insecticide aimed at killing the corn borer pest.

The committee said in a report that the research was not sufficient to warrant a ban, particularly as a previous field-based study showed that the modified maize had "no measurable effects" on lacewings.

But the government may come under renewed pressure to bring in a moratorium on the commercialization of genetically modified crops following preliminary findings from the Rowett Research Institute of nutrition in Aberdeen, Scotland. These found that rats fed on modified potatoes suffer damage to their immune systems.

Gulf War syndrome 'not due to depleted uranium'

[WASHINGTON] The US Department of Defense last week released a report concluding that it could find no evidence that depleted uranium caused or is causing the ill-defined collection of illnesses known as Gulf War syndrome.

"The bottom-line conclusion, based on a comprehensive review of available data and a science-based methodology, is that exposures to depleted uranium's heavy metal (chemical) toxicity or low-level radiation are not a cause of the undiagnosed illnesses afflicting some Gulf War veterans," the report says.

Depleted uranium is a by-product of the production of nuclear fuel or weapons.

Further evidence in trial over blood scandal

[TOKYO] A tape recording has revealed that Japanese government officials knew in 1983 that unheated blood products used to treat Japan's 5,000 haemophiliacs may have been tainted with HIV.

The tape was played last month during the trial of a former health ministry official charged with professional negligence for failing to order the withdrawal of contaminated blood products despite knowing the risks of HIV infection.

Some 1,800 haemophiliacs have contracted HIV through using the blood products (see *Nature* 379, 663; 1996).

Takeshi Abe, former vice-president of Teikyo University Medical School, who headed a government AIDS study panel, is heard on the tape saying that he feels as though he is injecting poison into his patients every day.

Rockefeller cash to bring fresh expertise to WHO

[MUNICH] The World Health Organization (WHO) and the Rockefeller Foundation have created a US\$2.5 million fund aimed at recruiting leading experts to the WHO.

The Global Health Leadership Fund is limited to two years and is designed to support reform of the WHO spearheaded by

Gro Harlem Brundtland, the organization's new president. Brundtland is keen to create partnerships with other United Nations agencies, governmental and non-governmental organizations, and with the research community.

SOHO satellite could be out of action for weeks

[WASHINGTON] Ground controllers have re-established limited radio contact with the tumbling SOHO Sun-observing satellite, but say it could be weeks before they know whether the spacecraft can resume science operations.

Although SOHO answered a radio signal on 3 August after six weeks of silence, onboard power is still low. The hydrazine fuel that powers the satellite's thrusters is presumed frozen, leaving the spacecraft unable to turn its solar arrays towards the Sun to recharge its batteries.

Project managers still hope that power will be slowly restored as the orientation of the arrays changes over the next six weeks.

France gives go-ahead for collider tunnelling

[LONDON] France's prime minister Lionel Jospin has given the green light for construction work to begin on the Large Hadron Collider, a particle accelerator being built on the French-Swiss border. The Swiss authorities gave approval for tunnelling work to begin earlier this year.

French approval came after "a long and painstaking environmental impact study", according to the European Laboratory for Particle Physics (CERN) in Geneva, which is building the collider.

Lewinsky controversy hits energy department

[WASHINGTON] The appointment of Bill Richardson as US energy secretary could be delayed again following newspaper allegations concerning the circumstances in which the former US ambassador offered a job to Monica Lewinsky, the celebrated former White House intern.

A newspaper alleged last week that Richardson created a junior public affairs position for Lewinsky at the United Nations in New York, after being encouraged to do so by the White House when Lewinsky's alleged relationship with President Bill Clinton first came under investigation. Lewinsky was offered the position, but turned it down.

At a hearing before the Senate Energy Committee, Richardson said the job already existed. The committee chairman has asked Clinton not to swear in Richardson until the allegation that he lied at the hearing has been investigated.

French reforms should go further

Sir — Your coverage of the proposed reforms of French university and medical research (INSERM) and, in particular, the resistance of the universities, INSERM and CNRS to the changes suggested by Claude Allègre, the Minister of Education and Research, while accurate so far as it goes, does not mention the main factor at stake (*Nature* 393, 97, 102 & 106; 1998).

The main thrust of the reforms proposed by Allègre is obviously the attempt to break, or at least reduce in French academia, the power of the corporate system, which has evolved naturally, as in other (particularly Latin) countries, from the family to the clan and finally the professional interest groups. *Les grandes écoles* are just the tip of an iceberg of overlapping circles of influence at all levels that, as in many other professions, are an overwhelming preoccupation of French scientists and professors.

At a recent meeting for the graduating class of one of the main *lycées* of Paris, two students of biology were among the professionals invited to be questioned about careers. One student had been at a *grande école* and his classmate at a university. The former was full of assurance and the latter slightly apologetic, but both agreed that being at a *grande école* increases by a factor of about ten the chance of being 'selected' at the Pasteur Institute. 'Grande école' not only means being taught by an elite of professors, after up to three years of painstaking

preparative classes to pass admission exams, but also is the chance to win the race for a place that gives security for life: a salary up to the end of studies and thereafter the protection of the 'old-timer' club.

Before a student knows how to hold a pipette, the future chemist or biologist knows that he or she has to choose the strongest possible 'school' or 'patron' to protect his or her place in the corporate system. This may end up taking at least 30% of a scientist's time, and up to 80% of time (including teaching) for a professor who may have to pay back many accumulated debts of assistance received.

The 'democratic' *commissions de spécialistes* that govern academic research in France are indeed elected in part, but of course, in addition to giving loyal service to the community, they also enable the corporate system to work in a political sense. Meeting in Paris, the same people distribute nationwide positions and money on an annual basis, and also judge research, with no direct input from independent or foreign experts. In the worst case, this leads to in-house intellectual censorship which is why, in the life sciences at least, much of the originality of research is crippled. Since there is no time for in-depth analysis, anyone with an unconventional idea is easily considered a pretentious crackpot — unless supported by a 'patron' accepted within the system. Originality and international recognition have little value within the corporate system.

Another handicap of French science is that hundreds of the best scientists have to devote endless hours to evaluating the others, a most serious social game — one's students, colleagues and possibly one's own careers are at stake. It is no wonder that Allègre, wishing to break the impact of networks carefully built up over so many years and instead have scientists competing solely by the criteria of international scientific and economy-oriented standards, has met so much resistance.

Of course *gradual* change is necessary, but attacking just *les grandes écoles* is only paying lip-service to the problem. Put bluntly, in France, the evaluation of academic and scientific achievement should be strictly separated from the distribution of positions and funds; and the biannual auto-evaluation of scientific achievement by a 'commission' should be replaced by site visits every three to four years, allowing enough time (and means) to justify a project without interference, and done by independent specialists, at least 30% of whom should be from outside France. Some successful attempts at such a system are already functioning, for example the ATIPE boards of the CNRS, which fund young laboratory leaders.

Klaus Scherrer

(Directeur de Recherche, CNRS, Paris)
Institut Jacques Monod, 2 Place Jussieu,
F-75251 Paris Cedex 05, France
e-mail: scherrer@ijm.jussieu.fr

Deceptive appearance

Sir — Hesse-Honegger's handwork is, as an aesthetic indicator of environmental pollution as described in Martin Kemp's Art and Science article, visually very attractive but may leave "corners where dark questions lurk" in areas other than those he mentions (*Nature* 392, 555; 1998).

The near-field of Chernobyl, Three Mile Island, and the "Chernobyl trails" in Sweden and Switzerland have very little in common radiologically, with exposures ranging from a fraction of natural levels to a thousandfold of normal. Finding consistent teratogenic effects in organisms generally known for low radiation sensitivity may have to do with poor control of important environmental parameters such as temperature during organ formation in these species. Twelve years after Chernobyl, the merits of such an assay on "radiation-mutated" bugs could have been scrutinized easily by blinding, that is, by separating bug collection and morphological assessment.

In the absence of controls, this kind of

superficially convincing science may suggest that the Earth is flat, or produce even less benign claims. Aesthetic hypotheses and even politics should similarly not be absolved from critical analysis, otherwise one is not obliged to accord them any importance outside the realm of aesthetics. I hope *Nature* will maintain its distinction between good and bad science; there is no room in between for nice (-looking) science.

Werner Burkart

BfS/Institute for Radiation Hygiene,
Ingolstaedter Landstr. 1,
D-85764 Oberschleißheim, Germany

East Germans succeed

Sir — I was disappointed by the portrayal of east German scientists in your recent article on neuroscience research in Magdeburg (see *Nature* 393, 725; 1998). Having worked as a doctoral student in the Institute for Neurobiology (IfN) in Magdeburg between January 1993 and November 1996 — the

first non-German westerner to do so — I feel well placed to comment.

You give the impression that, during the communist period, all east German scientists were either working for the Stasi secret police, or were uncreative or unproductive. But the publication records of IfN's departments during the five years after the fall of the Berlin Wall reveal that the department with the most impressive record was the one led by an east German; it contained several east German scientists. The most successful research group in the institute was led by an east German.

I would also like to point out that the only article produced by the IfN that has appeared in *Nature* was co-authored by an east German (Frey and Morris, *Nature* 385, 533–536; 1997), and that east German doctoral students compare very favourably in their grades for their theses with the "young, self-confident west German graduate students" that you describe.

Ritchie Brown

Physiology II, Heinrich-Heine-Universität,
POB 101007, D-40001 Düsseldorf, Germany

Falling satellites, rising temperatures?

Dian J. Gaffen

Satellite-based estimates of trends in atmospheric temperature have disagreed with estimates derived from other measurements. By taking into account declines in the orbital height of satellites, re-examination of those satellite data yields results that are in better accord with independent evidence of global warming.

There is no single, definitive, long-term record of global atmospheric temperatures. So, in their attempt to detect evidence of global warming (or otherwise), climatologists must try to make sense of temperature-trend results from a variety of imperfect measurement systems. In 1995, the Intergovernmental Panel on Climate Change reported¹ agreement among a group of datasets (including weather observations at the surface and from balloons, and data on sea ice, mountain glaciers and underground temperatures) showing that, over the past century, near-surface atmospheric temperatures have risen.

For the past 19 years, temperature measurements have also been made from space. For the lower part of the troposphere (the troposphere is the region of the atmosphere extending from the ground to about 10–15 km), these satellite-derived temperatures have, by contrast, shown slight cooling. This discrepancy has been controversial in the climate research community, and often features in broader discussions about environmental policy. On page 661 of this issue², Wentz and Schabel now present a correction to the satellite data that changes the lower-tropospheric trend from cooling to warming.

The main deficiencies of the *in situ* measurements made at surface stations and from balloons are the uneven sampling of the global atmosphere and changes in temperature-measurement systems over time¹. Polar-orbiting satellites largely overcome the global sampling problem by making measurements over most of the planet. Since 1979, a series of eight different NOAA (National Oceanic and Atmospheric Administration) polar-orbiters have carried microwave sounding units (MSUs) that measure the radiation emitted by atmospheric oxygen near a frequency of 60 GHz. Changes in the strength of this radiation can be attributed to changes in the internal energy of the oxygen molecule, which is directly related to the temperature of the atmosphere itself.

By combining MSU observations from

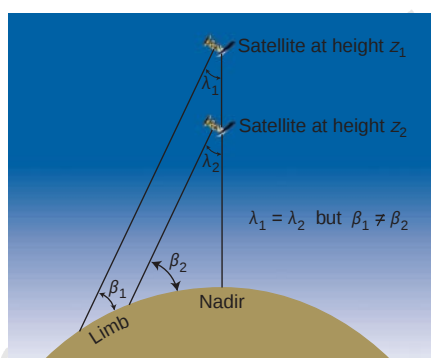


Figure 1 Angles on temperature trends in Earth's atmosphere (not to scale). A change in the height (z) of polar-orbiting satellites has little influence on near-nadir measurements of temperature. But as instruments scan away from a nadir, the interpretation of near-limb measurements (at angle λ) requires knowledge of the angle β , as described by Wentz and Schabel². Incorporating changes in z (and thus in β) in data from satellite-borne microwave sounding units, Wentz and Schabel have derived new estimates of temperature trends in the lower troposphere which indicate warming rather than cooling.

the NOAA satellites, Spencer and Christy³ developed the first satellite-derived dataset of global atmospheric temperatures. Temperatures in a relatively thick layer, mostly in the mid-troposphere but including part of the lower stratosphere (the atmospheric region above the troposphere), are based on MSU radiation measurements obtained near nadir, when the instrument is looking more or less straight down. Temperatures in a lower layer are derived using MSU data from both near-nadir and off-nadir (near-limb) observations, as shown in Fig. 1. Because the near-limb observations are more sensitive to temperature in the upper troposphere and lower stratosphere, subtracting near-limb observations from near-nadir measurements yields an estimate more representative of the lower troposphere.

Unexpectedly, Spencer and Christy's analysis revealed different trends in these

two layers, with the mid-troposphere showing warming and the lower troposphere showing cooling. Because this result ran counter to expectations about the nature of variations in global climate, and because the other records indicated warming over this period, researchers naturally looked for flaws in the analysis. Among the suggestions offered are that the data from different satellites are inconsistent^{4,5} or that they are contaminated by microwave emissions from solid and liquid water in the atmosphere^{6,7}. Nonetheless, the proponents of the MSU results have stood by their conclusions^{8,9}.

Wentz and Schabel² have brought to light a neglected aspect of the MSU data that influences the inferred lower-tropospheric temperature trends. The height of a satellite's orbit decreases slightly over time because of the drag of the atmosphere on the satellite. This effect increases during periods of high solar activity, because increased solar ultraviolet radiation warms the upper atmosphere and there are more frequent collisions of air molecules with the satellite (see Figs 1 and 2 on pages 662 and 663, respectively). As the satellite falls, the near-nadir measurements are not much affected. But the near-limb measurements are more seriously influenced because of the change in the MSU's angular view of the surface (Fig. 1). Spencer and Christy used periods of data overlap to merge data from sequential satellites, and this technique may have compounded the effects of orbital decay on data from each successive satellite.

Correction for this 'falling-satellite' effect brings the trends in the lower and mid-troposphere into much closer agreement. In fact, the correction term developed by Wentz and Schabel dominates the new lower-tropospheric trend for 1979–95. Before the correction, the MSU data showed the mid-troposphere to be warming at 0.03 °C per decade and the lower troposphere to be cooling at 0.05 °C per decade. After correction for the orbital decay, the lower troposphere shows warming of 0.07 °C per decade. Although this trend seems small, it is in line with longer-term estimates from surface data¹.

Before making too much of these new indications of warming, we should recognize some limitations in the satellite record. The trends presented cover a period of less than two decades and, given the variability of the global atmosphere and the multiplicity of potential sources of error, we cannot accurately estimate their confidence intervals. Wentz and Schabel estimate the lower tropospheric trend confidence interval as at least ± 0.05 °C per decade, at the 80% confidence level, which is much lower than the more commonly used 95% or 99% confidence level, and the trends in question are quite possibly not statistically significant at those higher levels. Viewed in this light, the discrepancy between cooling of 0.05 °C per

decade and warming of 0.07 °C per decade may be less notable than it appears.

Furthermore, the mere fact that inclusion of the effects of satellite orbital decay on the trend can completely change its sign makes one wonder what other factors may be influencing the apparent trends. Assiduous analysis of both the satellite and *in situ* datasets has revealed progressively more about the error characteristics of the data and their trends; nevertheless, much remains to be done to resolve differences among them.

The crux of the matter is that climatologists are relying on systems that were never designed for climate monitoring. Various groups have advocated improved observational networks^{10,11}, but they have yet to be realized. Complementary satellite and *in situ* observations could provide the global coverage and vertical resolution needed for reliable estimation of temperature trends in the troposphere and stratosphere.

Lessons from the analyses of MSU data can be applied to the development of a satellite-based climate monitoring system — obtaining the data needed for credible estimates of temperature trends requires attention to instrument calibration, continuous monitoring of the satellites for potential drift in the observations, sufficient overlap between successive satellites, and a commitment to a long-term programme of observations. *In situ* observations could be obtained from a radiosonde (weather balloon) network dedicated to climate research. Weather services around the world currently launch various types of expendable, and therefore relatively inexpensive, instruments to obtain temperature profiles for weather forecasting, and the types of instruments change frequently. A radiosonde network for climate monitoring could employ a standardized instrument with better accuracy and precision.

In short, we have to decide if we are really interested in monitoring the health of the Earth's atmosphere. If we are, more sophisticated arrangements for tracking the planet's temperature are needed.

Dian J. Gaffen is in the NOAA Air Resources Laboratory, 1315 East West Highway, Silver Spring, Maryland 20910, USA.

e-mail: dian.gaffen@noaa.gov

- Nicholls, N. et al. *The IPCC Second Scientific Assessment* (eds Houghton, J. T. et al.) 133–192 (Cambridge Univ. Press, 1996).
- Wentz, F. J. & Schabel, M. *Nature* **394**, 661–664 (1998).
- Spencer, R. W. & Christy, J. R. *Science* **247**, 1558–1562 (1990).
- Hurrell, J. W. & Trenberth, K. E. *Nature* **386**, 164–167 (1997).
- Prabhakara, C., Iacovazzi, R. Jr, Yoo, J.-M. & Dalu, G. *Geophys. Res. Lett.* **25**, 1927–1930 (1998).
- Hansen, J. et al. *Clim. Change* **30**, 103–117 (1995).
- Prabhakara, C., Nucciarone, J. J., Yoo, J.-M. & Dalu, G. *Clim. Change* **30**, 349–366 (1995).
- Christy, J. R. & Spencer, R. W. *Clim. Change* **30**, 97–102 (1995).
- Christy, J. R., Spencer, R. W. & Braswell, W. D. *Nature* **389**, 342 (1997).
- Karl, T. R. (ed.) *Long-Term Climate Monitoring by the Global Climate Observing System* (Kluwer, Dordrecht, 1996).
- Hansen, J. et al. *Clim. Change* **31**, 247–271 (1995).

Cancer

Awakening angels

David Lane

When normal mammalian cells are subjected to stress signals — oxygen deficiency, radiation, DNA damage or chemotherapeutic drugs, for example — the 'guardian of the genome', p53, is woken up. Then, p53-dependent gene transcription is increased and ubiquitin-dependent degradation of the protein is blocked, all of which leads to the p53-mediated induction of programmed cell death and/or cell-cycle arrest. But how do such stress signals awaken p53? One idea is that stress-activated protein kinases modify p53, protecting it from degradation and activating its function as a transcription factor. In support of this theory, many phosphorylated forms of p53 are found in cells. Moreover, p53 can exist in a latent state that cannot bind DNA, and it can be released

from this state by phosphorylation.

On page 700 of this issue, Woo *et al.*¹ present striking evidence that the DNA-dependent protein kinase (DNA-PK) is required for the p53 response to occur (Fig. 1a). This enzyme has long been an attractive candidate because it is activated by DNA damage, and p53 was one of its first substrates to be described. Furthermore, DNA-PK modifies the critical amino-terminal region, which controls the interaction of p53 with the transcriptional apparatus and with MDM2 (a protein that regulates p53)².

Enthusiasm for DNA-PK has, until now, been tempered by the finding that mice with severe combined immunodeficiency (SCID), which have a genetic defect in DNA-PK, can nevertheless respond to DNA damage in a p53-dependent manner^{3,4}. Thus, attention shifted to other related protein kinases that can either directly or indirectly affect phosphorylation of p53. These include the ATM protein, which is defective in patients with ataxia telangiectasia (AT)^{5,6}, and related ATR kinases.

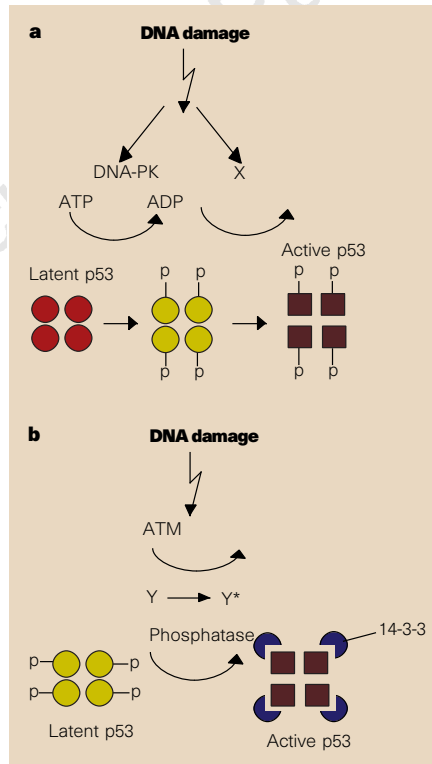


Figure 1 Possible ways to activate the p53 response. a, p53 is made in an inactive latent state. Woo *et al.*¹ have shown that the DNA-dependent protein kinase (DNA-PK) is needed to activate DNA binding and transcription by p53, and another factor (X) from nuclear extracts of irradiated cells is also required. b, p53 is made in an inactive latent state. It is phosphorylated but remains inactive until DNA damage activates the ATM kinase. ATM activates a p53-specific phosphatase (Y) that removes phosphate from serine 376, allowing 14-3-3 proteins to bind to the carboxy terminus of p53 and activate DNA binding and transcription.

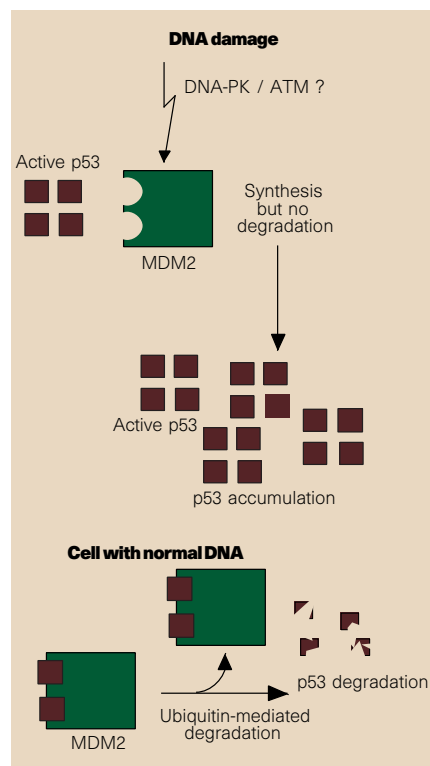


Figure 2 Another model for activation of p53. p53 is active in normal cells, but at a low level because it is constantly targeted for inactivation and degradation by MDM2. DNA damage blocks MDM2 or other components of the p53-degradation pathway, so that high levels of active p53 can now accumulate and, at this high concentration, may become phosphorylated.

An elegant alternative model to explain the action of these kinases has just been proposed (Fig. 1b). Here, the p53 response is initiated after DNA damage by the ATM-dependent activation of a phosphatase that removes a phosphate group from serine 376 of p53. This allows 14-3-3 proteins to bind to the carboxy terminus of p53, thus activating DNA binding and transcription by p53. Yet both of these models have been called into doubt by mutagenesis studies on p53. If all known phosphorylation sites — including those that are targets for DNA-PK and ATM — are mutated, p53 can still be stabilized and activated in response to stress. Thus, although p53 is clearly phosphorylated after DNA damage, the significance of this for the p53 response has been in doubt.

Another model for activation of p53 has arisen from an understanding of the critical role played by MDM2 (Fig. 2). This protein can target p53 for nuclear export and degradation^{7,8}. Disrupting the p53-MDM2 interaction in normal cells results in the accumulation of p53 and activation of p53-dependent transcription^{9,10}. The MDM2 gene is, itself, activated for transcription by p53, so this model implies that p53 is constitutively active, driving transcription of the protein that targets its own degradation. All that would be required to activate the p53 response would be to block this degradation pathway. In other words, MDM2 could be the sole target for a p53-activating stress signal. This model is satisfying because it explains why mutant p53 proteins are stable — when p53 is inactive, the cell lacks the MDM2 protein that is required for p53 degradation. In the most extreme version of this model, p53 is phosphorylated in DNA-damaged cells just because it is stable, has accumulated to high levels, and is therefore a better substrate for kinases.

How does the paper by Woo *et al.*¹ integrate with these radically different views of the response? Quite simply, it doesn't. First, the new work questions the penetrance of the SCID-mouse phenotype, by showing that cells from these animals retain residual DNA-PK activity. So DNA-PK may be needed for the p53 response after all. Second, the authors describe a completely DNA-PK-deficient mouse cell line (SCGR11) which, in response to DNA damage, can accumulate a form of p53 that cannot bind DNA and is not transcriptionally active.

At face value this does not fit with the MDM2 signal model. Because SCGR11 cells lack active p53, they should not contain any MDM2. So, their p53 should be constitutively stable (although perhaps all is not lost, as the authors do state that these cells contain constitutively high levels of p53). It is possible that, in activating the p53 response, the target for DNA-PK or the ATM/ATR kinases is not p53 itself, but MDM2 or another protein⁶. Indeed, if MDM2 is phosphorylated

by DNA-PK it can no longer bind to p53 (ref. 11). But again, Woo and colleagues' data do not support this. The authors favour the idea that the DNA-binding function of latent p53 is directly activated by DNA-PK, because they can only activate the DNA-binding activity of p53 *in vitro* with nuclear extracts of cells that contain DNA-PK. This cannot be done using pure DNA-PK — it needs to be complemented by another factor that is present in nuclear extracts of irradiated cells. This factor is activated even in cells that lack DNA-PK.

How will these findings be reconciled? It is already clear that many signalling pathways control the activation of p53, because cells that lack ATM show good p53-dependent responses to ultraviolet light but not to ionizing radiation¹². The p53 response is finely poised — just one double-strand break may be enough to trigger it, and responses to minute doses of ionizing radiation have been detected. Remarkably, genodose effects are detectable, and reduction to homozygosity is enough to compromise the response^{13,14}. T cells from the thymus of +/- p53 mice are relatively radioresistant compared with those from +/+ mice, challenging the simplest model of Knudsen's two-hit hypothesis for loss of tumour-suppressor genes. Delicate methods will be needed to

work out which pathways are physiological, and in which cell types, as tissue variation is evident.

It is to be hoped that, unlike the mediaeval theologians who disputed the number of angels that can occupy the point of a pin, the current controversy over the pathways that lead to the activation of p53 will lead to a genuine knowledge of how to awaken our 'guardian angel' gene. □

David Lane is at the Cancer Research Campaign Laboratories, Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK. e-mail: d.p.lane@dundee.ac.uk

1. Woo, R. A., McLure, K. G., Lees-Miller, S. P., Rancourt, E. E. & Lee, P. W. K. *Nature* **394**, 700–704 (1998).
2. Lees-Miller, S. P., Sakaguchi, K., Ulrich, S., Appella, E. & Anderson, C. W. *Mol. Cell. Biol.* **12**, 5041–5059 (1992).
3. Araki, R. *et al. Proc. Natl Acad. Sci. USA* **94**, 2438–2443 (1997).
4. Gurley, K. E. & Kemp, C. J. *Carcinogenesis* **17**, 2537–2542 (1996).
5. Kastan, M. B. *et al. Cell* **71**, 587–597 (1992).
6. Waterman, M. J. F., Stravadi, E. S., Waterman, J. F. L. & Halazonetis, T. D. *Nature Genet.* **19**, 175–178 (1998).
7. Haupt, Y., Maya, R., Kazan, A. & Oren, M. *Nature* **387**, 296–299 (1997).
8. Kubbutat, M. H., Jones, S. N. & Vousden, K. H. *Nature* **387**, 299–303 (1997).
9. Blaydes, J. P. *et al. Oncogene* **14**, 1859–1868 (1997).
10. Bottger, A. *et al. Curr. Biol.* **7**, 860–869 (1997).
11. Mayo, L. D. *et al. Cancer Res.* **57**, 5013–5016 (1997).
12. Khanna, K. K. & Lavie, M. F. *Oncogene* **8**, 3307–3312 (1993).
13. Clark, A. R. *et al. Nature* **362**, 849–852 (1993).
14. Lowe, S. W. *et al. Nature* **362**, 847–849 (1993).

Plasma physics

Surfing the wake

Robert Bingham

For almost 60 years, physicists have used strong, large-scale electric fields to accelerate charged subatomic particles. Unfortunately, these conventional particle accelerators, RF linacs, have reached a limit: at an electric field strength of about 10^7 V m^{-1} , the support structures break down — sparks fly, short-circuiting the accelerator. So, in order to reach higher particle energies one is forced to build ever larger linear accelerators, or bend the particles into a circular path using magnets (which produces another problem, as the particles then radiate away much of their energy). In theory, this electric-field problem can be overcome by plasma particle accelerators, because plasma, as an ionized medium, has already broken down and is therefore not susceptible to electron dissociation. Plasma waves can generate accelerating electric fields thousands of times greater than the most powerful conventional accelerators. And in *Physical Review Letters*, Amiranoff and colleagues now report¹ an experiment that brings the 'table-top' plasma accelerator a step closer to reality.

Amiranoff *et al.*¹ show that powerful accelerating gradients and high electron

energies can be achieved by the laser wake-field acceleration scheme, one of several plasma-based processes being pursued. Their results swiftly follow the report by Gordon *et al.*² of 94-MeV electrons produced by a related process — self-modulated laser-wakefield acceleration. These experiments, together with the plasma-beat-wave results³, demonstrate the potential of plasma accelerators.

The laser-wakefield accelerator was first proposed by Tajima and Dawson in 1979⁴, together with the laser-beat-wave accelerator, but it has only been realized now because it required the development of terawatt lasers, which produce sub-picosecond pulses with the necessary high intensity ($\geq 10^{18} \text{ W cm}^{-2}$).

The new generation of high-intensity lasers can be focused to produce enormous electric field strengths — the electric field of a $10^{18} \text{ W cm}^{-2}$ laser is $3 \times 10^{12} \text{ V m}^{-1}$, comparable to the field in atoms. But used directly, the laser field is not effective at particle acceleration. As the electric field of the laser is perpendicular to the propagation direction, the maximum gain is limited by the distance the particle moves across the wavefront before the electric field changes sign.

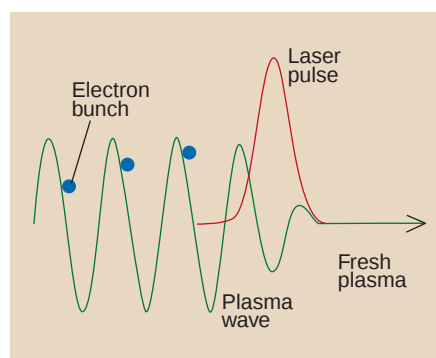


Figure 1 A laser-wakefield accelerator. A sub-picosecond laser pulse of ultra-high intensity ($\geq 10^{18} \text{ W cm}^{-2}$) excites a wake of plasma oscillations moving at close to the speed of light. Electrons can gain energy from this wave by 'surfing' down its electric-field gradient.

In the longer term, the electron oscillates up and down in the laser field, gaining no energy on average.

By using a plasma into which laser energy can be coupled, however, we can generate an electron plasma wave moving in the direction of the laser, whose electric field is parallel to the direction of propagation. In contrast to earlier laser-beat-wave experiments³, which used a two-frequency laser to excite the plasma wave, laser-wakefield acceleration uses a single, short laser pulse to excite a wake of electron plasma waves that trails behind the laser pulse, much like the wake of a motor boat (Fig. 1). Plasma electrons are depleted in the region of the laser pulse, pushed out by its radiation pressure. That leaves behind a region of net positive charge, which then pulls electrons back in, setting up an oscillation.

The acceleration process acts only on electrons moving at close to the phase velocity of the plasma wave. They are accelerated by the wave's electric field until they eventually outrun it, rather like a surfer on an ocean wave. For a high energy gain, the trick is to generate a coherent, large-amplitude electron plasma wave moving close to the speed of light over a sizeable distance. To do this, the laser pulse must be brief — if the laser pulse length is long compared with the natural period of the plasma wave, then the plasma wave energy is re-absorbed by the trailing part of the laser pulse.

In the latest experiment¹, a single-pulse neodymium laser and a 3-MeV electron beam are focused to the same point. To achieve optimal wave generation, the laser pulse lasts 400 fs, matching the electron-plasma-wave period. The intensity of the laser is $4 \times 10^{17} \text{ W cm}^{-2}$ and is roughly constant over a length of about 2 mm, setting the size of the acceleration region. The measured maximum energy gain is 1.6 MeV, corresponding to a maximum longitudinal accelerating electric field of about 1.5 GV m^{-1} —

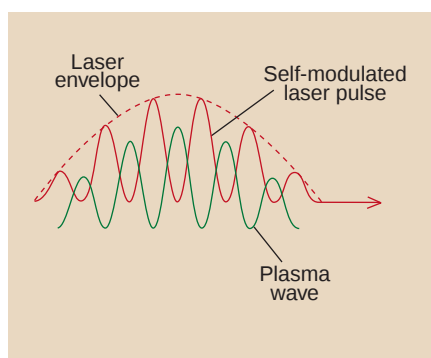


Figure 2 A self-modulated laser-wakefield accelerator. This hybrid scheme exploits an instability known as stimulated Raman forward scattering. The sharp leading edge of a long laser pulse creates a region of low electron density. The trailing part of the laser pulse scatters from this density perturbation, producing a modulated laser wavepacket that breaks the laser pulse into a series of shorter pulses, each helping to generate the plasma wave.

more than a hundred times the conventional accelerator limit, and 10% of the theoretical maximum in their plasma.

The laser wakefield has a number of advantages over other processes. As the electron plasma wave's maximum amplitude is immediately behind the pulse, it requires no growth phase as do other schemes such as the laser beat wave³. In the laser-beat-wave accelerator, two laser beams have a frequency difference precisely equal to the plasma frequency. As the plasma wave grows to large amplitude, the oscillating electrons that form the wave reach relativistic velocities, which decreases the plasma frequency, resulting in a de-tuning between the wave

and the laser beat frequency. Furthermore, if the ion plasma period is shorter than the time for the plasma wave to grow, ions produce instabilities and turbulence. These effects all limit particle acceleration.

Another rival process is self-modulated laser-wakefield acceleration (Fig. 2), which Gordon *et al.*² have used to generate electrons of nearly 100 MeV, in a distance of several millimetres. But this process may not be the best candidate for very high energy accelerators, because it relies on an instability to create the electron plasma wave, and so there is little control over the final amplitude of the wave, which can break and so produce the highest-energy electrons.

The new result completes a set of experiments proving the principle of plasma acceleration. But to do interesting experiments, we need energies greater than 1 GeV, with good beam quality and particle numbers of 10^8 to 10^{10} . That will require acceleration lengths of tens of centimetres. In present-day experiments, diffraction limits this to a few millimetres, but this could be extended using plasma channels, in which the radiation is trapped either by self-focusing or by a refractive-index gradient. It should also be possible to use higher plasma densities in order to increase the accelerating electric field, and to have a series of devices accelerating particles in several stages. □

Robert Bingham is at the Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK.

e-mail: r.bingham@rl.ac.uk

1. Amiranoff, F. *et al.* *Phys. Rev. Lett.* **81**, 995–998 (1998).
2. Gordon, D. *et al.* *Phys. Rev. Lett.* **80**, 2133–2136 (1998).
3. Everett, M. *et al.* *Nature* **368**, 527–529 (1994).
4. Tajima, T. & Dawson, J. M. *Phys. Rev. Lett.* **43**, 267–270 (1979).

Signal transduction

Marked for nuclear export?

Nullin Divecha

The nucleus of a eukaryotic cell contains all the information needed for that cell to proliferate, differentiate or undergo programmed death (apoptosis). Which of these programmes is followed depends on signals sent to the nucleus in response to external cues such as growth factors. These bind to receptors at the plasma membrane and transduce signals to the inside through the generation of second messengers. On page 697 of this issue, Topham *et al.*¹ present a new twist to one of the mechanisms by which this occurs.

Signal transduction into the nucleus can occur through the generation of diacylglycerol (DAG), a membrane-bound second messenger that activates protein kinase C (PKC)². Upon growth-factor stimulation, PKC can translocate into the nucleus³ and

phosphorylate a number of nuclear targets (such as the nuclear lamins⁴). Herein lies the problem. Some isoforms of PKC are absolutely dependent on DAG and, as this lipid is unlikely to translocate with the kinase, how are the levels of DAG in the nucleus coordinated with the presence of nuclear PKC? Topham and co-workers suggest that PKC can, itself, regulate the levels of nuclear DAG by inhibiting the nuclear entry of DAG kinase ζ (DGK- ζ), an enzyme that removes DAG. Furthermore, by genetically manipulating the levels of DAG kinase, the authors show that nuclear DAG is essential in regulating the cell cycle.

The story behind nuclear DAG and its regulation has been contentious⁵. Studies — which usually involve isolating nuclei before and after various stimuli, then measuring the

levels of nuclear DAG — have been met with cries of “contamination” (presumed to be non-nuclear membranes), and the most feared scientific expletive, “artefact”. But some studies have shown that certain growth factors can stimulate the production of DAG in the nucleus (with no changes at the plasma membrane), and that other growth factors can stimulate changes at the plasma membrane (with no changes in the nucleus). This suggests that signalling through DAG at these two membranes can be regulated independently^{6,7}.

Although data have accumulated in favour of a causal relationship between the regulation of nuclear DAG levels and cell proliferation⁵ or differentiation^{8,9}, no direct link has been shown. This is probably because there are very few specific ways to regulate the levels of nuclear DAG at will (neither chemical inhibitors, nor nuclear-specific enzymes that generate or metabolize DAG). But Topham and co-workers have now come close to getting around these problems.

Using immunofluorescence (and thus ruling out contamination by other organelles), the authors show that DGK- ζ — which was previously found to be localized in the nucleus¹⁰ — shuttles between the cytoplasm and the nucleus. Interestingly, the level of DGK- ζ in the nucleus is controlled by the PKC-mediated phosphorylation of a MARCKS (myristoylated alanine-rich C-kinase substrate) homology domain, which seems to encode a nuclear-localization signal.

The idea is that the phosphorylated form of DGK- ζ is excluded from the nucleus, leading to an increase in the levels of nuclear DAG — which the authors see after treatment with a growth factor (in this case, epidermal growth factor, EGF). It is not clear whether DGK- ζ is phosphorylated in the cytosol or in the nucleus (after translocation of PKC, which occurs in Swiss 3T3 cells in response to EGF³). But the outcome is a positive-feedback loop, such that PKC-mediated phosphorylation of DGK- ζ leads to an increase in the levels of nuclear DAG (Fig. 1).

So is there a direct link between nuclear DAG and the cell cycle? Topham and co-workers find that overexpression of wild-type DGK- ζ leads to decreased levels of nuclear DAG, with no effect on the levels in the plasma membrane. Overexpression also leads to an increased cell-doubling time, and the cells accumulate in the G0/G1 stage of the cell cycle. Because the authors found what controls the nuclear translocation of DGK- ζ , they were able to perform the ultimate control experiment — expression of an active enzyme that lacked the nuclear-localization sequence.

As with all interesting papers, this one creates more questions than it answers. Is this a widespread mechanism for the regula-

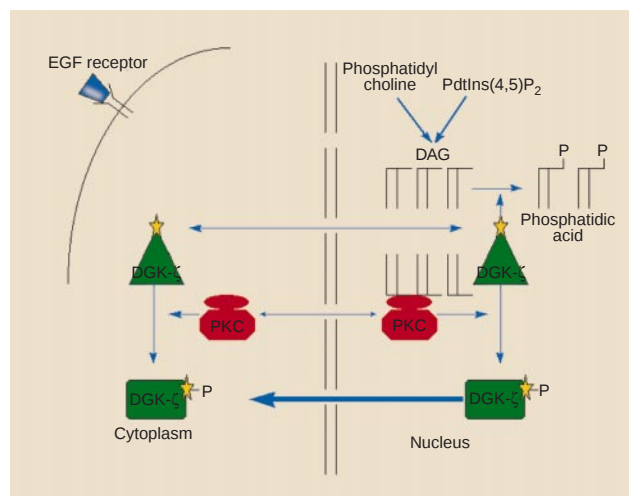


Figure 1 Possible mechanisms for the regulation of nuclear diacylglycerol (DAG), based on the results of Topham *et al.*¹ and others. DAG kinase ζ (DGK- ζ) phosphorylates DAG to phosphatidic acid, thereby preventing DAG from activating protein kinase C (PKC). But PKC can phosphorylate DGK- ζ at its MARCKS site (yellow star), perhaps in response to stimulation by the epidermal growth factor (EGF).

tion of nuclear DAG? Does phosphorylation of DGK- ζ lead to increased export or decreased import of this kinase? And is translocation of PKC to the nucleus required for the increase in nuclear DAG? Data from my own, and other, labs¹¹ indicate that PKC is not required for increased nuclear DAG levels in response to some growth factors. Stimulation by insulin-like growth factor-I, for example, has been shown^{7,12} to result in the activation of a phosphoinositidase C. So, levels of nuclear DAG may be regulated by many mechanisms. Moreover, it has been suggested that nuclear DAG is also required for the regulation of nuclear PKC at G2/M of the cell cycle^{13,14}. Because DAG can be generated in several ways (Fig. 1), leading to the production of structurally different DAGs¹⁵, it is possible that different pools of nuclear DAG coordinate the activation of PKC at different points in the cell cycle.

With the usual reservations, Topham *et al.* show that nuclear DAG may be essential in regulating the cell cycle. The next steps will be the molecular identification and charac-

terization of other enzymes involved in the metabolism of nuclear DAG. □

Nullin Divecha is in the Department of Cellular Biochemistry, The Netherlands Cancer Institute, Plesmanlaan 121, 1066cx Amsterdam, The Netherlands.

e-mail: ndivecha@nki.nl.

1. Topham, M. K. *et al.* *Nature* **394**, 697–700 (1998).
2. Nishizuka, Y. *Science* **258**, 607–614 (1992).
3. Neri, L. M. *et al.* *FEBS Lett.* **347**, 63–68 (1994).
4. Goss, V. L. *et al.* *J. Biol. Chem.* **269**, 19074–19080 (1994).
5. Divecha, N. *et al.* *Biochem. Soc. Trans.* **25**, 571–575 (1997).
6. Banfic, H., Zizak, M., Divecha, N. & Irvine, R. F. *Biochem. J.* **290**, 633–636 (1993).
7. Divecha, N., Banfic, H. & Irvine, R. F. *EMBO J.* **10**, 3207–3214 (1991).
8. Divecha, N., Letcher, A. J., Banfic, H. H., Rhee, S. G. & Irvine, R. F. *Biochem. J.* **312**, 63–67 (1995).
9. Martelli, A. M. *et al.* *Cell Signal.* **7**, 53–56 (1995).
10. Goto, K. & Kondo, H. *Proc. Natl Acad. Sci. USA* **93**, 11196–11201 (1996).
11. Martelli, A. M. *et al.* *Biochem. Biophys. Res. Commun.* **177**, 480–487 (1991).
12. Martelli, A. M. *et al.* *Nature* **358**, 242–245 (1992).
13. Sun, B., Murray, N. R. & Fields, A. P. *J. Biol. Chem.* **272**, 26313–26317 (1997).
14. Thompson, L. J. & Fields, A. P. *J. Biol. Chem.* **271**, 15045–15053 (1996).
15. Hodgkin, M. N. *et al.* *Trends Biochem. Sci.* **23**, 200–204 (1998).

Ecology and statistics

Nonlinear sheep in a noisy world

Nils Chr. Stenseth and Kung-Sik Chan

Population ecologists have long been interested in two key topics¹: the first is the relative importance of intrinsic factors (such as the inhibition of reproduction at high population densities) and extrinsic environmental variations in determining population fluctuations; the second is nonlinearity in the processes that generate these fluctuations. On page 674 of this issue, Grenfell and co-workers² discuss both of these issues. From their studies of fluctuations in two populations of Soay sheep living on islands 3.5 km apart in the St Kilda archipelago (in the Outer Hebrides), they provide a new insight into the synchronization

of population fluctuations at separate, but not too distant, locations^{3,4}.

Robert May's pioneering paper of 1974 showed that even very simple, deterministic population models can have complicated dynamics (including chaos, which is not easily distinguishable from a random pattern), provided the models are sufficiently nonlinear⁵. This work stimulated ecologists to look for the presence of nonlinear interactions (that is, situations in which two and two do not make four). Although ecologists generally appreciate the importance of external (abiotic) environmental variability, be it deterministic or

Modelling nonlinear interactions – the TAR approach

Let X_t be the logarithmic abundance (population size) at time t (as in Grenfell and colleagues' study, $X_t = \log(N_t)$, where N_t is the proper abundance at time t). Assuming only direct density dependence (that is, a reduction in the population growth rate owing to factors such as inhibition of reproduction at high population densities), a TAR model (of the SETAR type) is given as:

$$X_t = \begin{cases} a_{1,0} + (1 - a_{1,1})X_{t-1} + \epsilon^{(1)}_t & X_{t-1} \leq \theta \\ a_{2,0} + (1 - a_{2,1})X_{t-1} + \epsilon^{(2)}_t & X_{t-1} > \theta \end{cases}$$

Here, θ is a parameter defining two regimes; $a_{1,0}$ and $a_{2,0}$ are the maximal growth rates in the lower and the upper regime (that is, when the log abundance X_{t-1} is below and above the

threshold, respectively); $a_{1,1}$ and $a_{2,1}$ are the strengths of the direct density dependence in the two regimes; and $\epsilon^{(1)}_t$ and $\epsilon^{(2)}_t$ are the random components in the two regimes^{10,14}.

The MSS model⁸ discussed by Grenfell and co-workers⁷ half a decade ago is essentially a TAR model on a logarithmic scale. The basic MSS model is given as $N_t = N_{t-1}R / (1 + (N_{t-1}/K)^b)$, where N_t is the population size at time t , R is the maximal net population growth rate, b is a measure of the strength of the density-dependent reduction of the net population growth rate, and K is some proxy for the carrying capacity (that is, the maximum population size that can be sustained by the

By taking logs on both sides of the expression defining this model, we obtain the approximation:

$$X_t = \begin{cases} r + X_{t-1} & X_{t-1} \leq k \\ (r + b \cdot k) + (1 - b)X_{t-1} & X_{t-1} > k \end{cases}$$

where $r = \log(R)$, b and $k = \log(K)$ are ecologically interpretable parameters.

We could incorporate randomness – a main component of Grenfell and colleagues' argument – in the maximal net population growth rate (R), in the carrying capacity of the area (K), or both. Assuming that the environmental noise in the carrying capacity is fairly small (justified by Grenfell and colleagues' results), we may ignore

the randomness in the threshold, and write:

$$X_t = \begin{cases} r + X_{t-1} + \epsilon^{(1)}_t & X_{t-1} \leq k \\ (r + b \cdot k) + (1 - b)X_{t-1} + \epsilon^{(2)}_t & X_{t-1} > k \end{cases}$$

This is a constrained continuous TAR model¹⁶, because its five statistical parameters (one in front of X_{t-1} in the lower region as well) are expressed in terms of three ecological parameters (R , b and K). A continuous TAR model cannot give a better fit than a TAR model (as found by Grenfell *et al.*). However, its advantage lies in parsimony – we should not forget 'Occam's razor': "entities are not to be multiplied beyond necessity". So, a continuous TAR approach seems preferable from an ecological point of view. **N.C.S. & K.-S.C.**

random, most ecological models of population dynamics (deriving, in one way or another, from Lotka and Volterra) are essentially deterministic¹. Abiotic, random variability has been – and still is – emphasized much less. In other words, all population-dynamic variations are, by and large, seen as effects of deterministic processes.

But in their new study, Grenfell and colleagues' main point is that randomness is indispensable for ecological theory, particularly in a nonlinear setting. Specifically, they address how extrinsic environmental variability and nonlinearity interact. The authors show that, whereas the nonlinear population growth rate tends to magnify differences between the two sheep populations, common environmental factors (such as weather) tend to synchronize the populations. In other words, environmental processes need to be very similar to synchronize two populations.

Grenfell and co-workers applied a statistical approach to the nonlinear, time-series modelling that was originally developed by Howell Tong⁶. The authors used the self-exciting threshold autoregressive (SETAR) model, a sub-class of the nonlinear, statistical TAR models⁶. The SETAR model is roughly equivalent to an ecological model that Grenfell and co-workers themselves used earlier⁷ – the nonlinear model by Maynard Smith and Slatkin⁸ (MSS model; see box).

The TAR models are based on the principle of decomposing the state space (the mathematical space defined by the state vectors; in this case, the density of animals) into two or more regions, also called 'regimes'. Linear autoregressive models are then fitted to each regime. Such models can reproduce a wide spectrum of dynamical behaviour^{9,10}, including chaos. Tong and Lim¹¹ were the first to apply this class of model to the classic Canadian lynx data¹², which showed similar density fluctuations to the populations of Soay sheep now studied by Grenfell *et al.* But Grenfell and colleagues are among the first ecologists^{10,13} to apply such models.

To some ecologists the TAR models may seem rather crude representations of the underlying population-dynamic processes. But the TAR model used by Grenfell and colleagues provides a statistically practical representation of a commonly discussed ecological model. In fact, many nonlinear models may, on a log scale, be approximated by a statistical TAR model. The key thing to do is to log-transform the original models^{6,14,15}.

More work on the interface between ecology and statistics is needed to clarify issues relating to the modelling of nonlinear dynamics in a noisy world. Such work will help biologists to tackle conceptual issues, such as those discussed by Grenfell and colleagues, and allow statisticians to develop

useful methods that can be applied in fields beyond ecology and biology. Such developments are best undertaken by ecologists and statisticians interacting closely – just as in the multidisciplinary team led by Bryan Grenfell. □

Nils Chr. Stenseth is in the Division of Zoology, Department of Biology, University of Oslo, Blindern, N-0316 Oslo, Norway.
e-mail: n.c.stenseth@bio.uio.no.

Kung-Sik Chan is in the Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242, USA.
e-mail: kchan@stat.uiowa.edu.

1. Begon, M., Harper, J. & Townsend, C. R. *Ecology* (Blackwell, Oxford, 1996).
2. Grenfell, B. T. *et al.* *Nature* **394**, 674–677 (1998).
3. Moran, P. A. P. *Aust. J. Zool.* **1**, 291–298 (1953).
4. Ranta, E., Kaitala, V., Lindström, J. & Helle, E. *Oikos* **78**, 136–142 (1997).
5. May, R. *Science* **186**, 645–647 (1974).
6. Tong, H. *Non-linear Time Series* (Oxford Univ. Press/Clarendon, Oxford, 1990).
7. Grenfell, B. T. *et al.* *Nature* **355**, 823–826 (1992).
8. Maynard Smith, J. & Slatkin, M. *Ecology* **54**, 384–391 (1992).
9. Lim, K. S. *J. Time Ser. Anal.* **13**, 119–132 (1992).
10. Stenseth, N. C. *et al.* *Proc. R. Soc. Lond. B* (in the press).
11. Tong, H. & Lim, K. S. *J. R. Stat. Soc. B* **42**, 245–292 (1980).
12. Lin, T. C. & Pourahmadi, M. *JRSS C Appl. Stats* **47**, 187–201 (1998).
13. Framstad, E. *et al.* *Proc. R. Soc. Lond. B* **264**, 31–38 (1997).
14. Stenseth, N. C. *et al.* *Proc. R. Soc. Lond. B* **263**, 1423–1435 (1996).
15. Williamson, M. H. *The Analysis of Biological Populations* (Arnold, London, 1972).
16. Chan, K. S. & Tsay, R. S. *Biometrika* **85**, 413–426 (1998).

NATURE

100 YEARS AGO

I should be obliged by your inserting the following experience if you think it remarkable. Yesterday we killed an adder (?) here, about 38 inches long; and seeing that he had made a meal evidently some little time before, out of curiosity we opened him, and extracted a large toad, which was about half-way down the snake's interior, or about 18 inches. The toad, whose head was much wider than the snake's, and whose body was many times as large as his enemy's head, we of course all thought must be dead; and we laid him on a flowerbed, wondering how he could have got inside the snake at all, for it certainly seemed a case of the greater being contained in the less. Of course we knew the marvellous stretching powers of a snake's jaws, but this seemed to eclipse them all. As we watched the toad he seemed to move, so we bethought ourselves of trying to revive him, and, after pouring water freely over him, and whisky and water down his throat, we were intensely astonished to see him revive; so much so that he stood up on all-fours, blown out like a balloon, and made a kind of a dart at a stick in the most comical way. Eventually "Jonah," as we promptly christened him, disappeared amongst the flowers. Can any of your readers quote a like case of resuscitation? Perhaps some of them might be able to afford information as to the probable duration of the toad's entombment.

From *Nature* 11 August 1898.

50 YEARS AGO

For nearly eighty years, one of the most prominent items of anatomical evidence in support of the antiquity of man in Australia has been a fragment of a supposed human molar, found by Krefft in the Wellington Caves of New South Wales. ... Recently, the human origin of the fragment has been questioned by Dr. T. D. Campbell, and I have considered *de novo* its possible relation to the lower mammals. ... After direct comparison with relevant material, I have no doubt that Krefft's find was derived from the posterior half of the upper fourth premolar or seccator, of the right side, of *Macropus (Protemnodon) anak*, Owen, a giant Pleistocene 'wallaby'....

From *Nature* 14 August 1948.

Thermodynamics

Cool sounds

Peter T. Landsberg

The turbines and crankshafts of the mechanical engineers have proved to be very useful, as everyone knows. But they produce noise, and unfortunately need oiling every now and again. How much better if engines could be constructed whose moving parts have been eliminated! Solar cells, for the production of electricity, are one way to go: the result is an elegant 'engine' in which only the electrons are the moving parts. In a paper¹ in *Physical Review Letters*, Reid, Ward and Swift describe another, but less familiar, transition to engines without moving parts. Moreover, by using a single gas both as the working part of the engine and as a heat-exchange medium, an old theoretical efficiency limit can be transgressed.

The principle that they explore is embodied in the thermoacoustic thermal engine and the closely related thermoacoustic refrigerator². In a thermoacoustic heat engine, sound waves are generated by a heat flow, and the oscillatory motion of the gas in the sound wave is the only moving part in the thermoacoustic set-up.

The sound wave is contained in a resonator, closed at both ends. The engine receives heat at a high temperature from a heat source via a hot heat exchanger, and rejects heat at a lower temperature via a cold heat exchanger to a heat sink. In each acoustic cycle, the sound wave moves the gas back and forth a small fraction of the distance between these two heat exchangers, within a third heat exchanger called the 'stack'. The stack consists of a number of parallel plates open to the gas flow and of high heat capacity. Any given parcel of gas absorbs heat from one location in the stack during one extremum of its motion, and rejects heat to another, cooler location in the stack while near the other extremum of its motion. In this way, heat is ferried from the heat source to the heat sink.

Each such parcel of gas absorbs its heat while the sound wave imposes high pressure, and rejects its heat while the sound wave imposes low pressure. Hence the gas enjoys thermal expansion at high pressure and thermal contraction at low pressure, producing work and thereby reinforcing the sound wave.

As with many types of heat engine, the thermoacoustic engine can be reversed to become a refrigerator. Like the heat engine, the thermoacoustic refrigerator uses a standing wave in a resonator, but with all the energy flows reversed: the system receives heat at low temperature and rejects it at higher temperature, while absorbing acoustic power

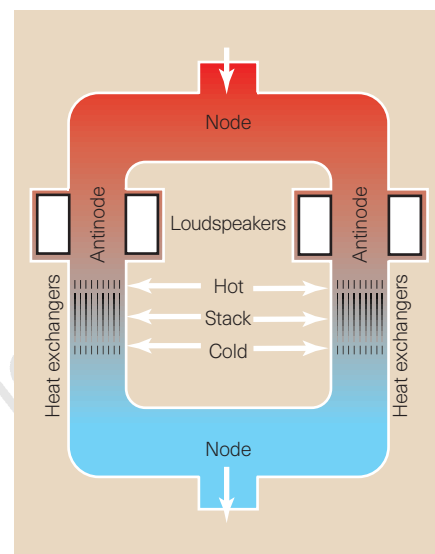


Figure 1 A thermoacoustic refrigerator with an externally applied flow. By using the gas stream to be cooled as part of the mechanism, such a device could be more efficient than a conventional closed-off thermoacoustic refrigerator.

supplied by loudspeakers. There is a theoretical lower limit (based on the Carnot cycle) to the work required to produce a desired cooling power, proportional to the temperature difference. In the special case that the cooling power is used to cool a gas stream from the refrigerator's own high temperature to its low temperature, the lower limit on the work required to process a given gas flow rate is proportional to the square of the temperature difference.

But in a new development¹, experimentally demonstrated at the Los Alamos National Laboratory, the resonator is opened at both ends, and the flowing gas stream to be cooled is used as the thermoacoustic working gas (Fig. 1). This leads to the elimination of the cold heat exchanger that previously carried heat from the gas being cooled to the thermoacoustic gas. Furthermore, by providing almost a continuum of staged refrigerators to the flowing gas stream, this arrangement reduces the thermodynamic lower bound on the necessary mechanical power to half of the Carnot lower bound mentioned above. In the working model that the authors have built, the actual power required is still high, at about 25 times the ordinary Carnot lower limit, but with some simple refinements (such as increasing the size and pressure to reduce viscous losses) improvements are undoubtedly possible.

This technique won't revolutionize

ordinary air conditioning, as that is based on cooling air that is already inside a building, rather than pumping in warm air from outside and cooling that. But in-stream thermoacoustic cooling may find other more specialized uses, such as dehumidifying a compressed air stream. □

Peter T. Landsberg is in the Faculty of Mathematical Studies, University of Southampton, Southampton SO9 5NH, UK.

e-mail: p.t.landsberg@soton.ac.uk

1. Reid, R. S., Ward, W. C. & Swift, G. W. *Phys. Rev. Lett.* **80**, 4617–4620 (1998).
2. Swift, G. W. *Physics Today* **48**, No. 7, 22–28 (1995).

Immunology

Burnet's unhappy hybrid

Klaus Rajewsky

The clonal-selection theory¹ proposed by Burnet some 40 years ago holds that antibody-forming cells (B lymphocytes) are precommitted — before contact with antigen — to a single antibody specificity. The B cells then express antibody of that specificity as an antigen receptor on their surface. This theory was borne out by molecular analysis showing that individual members of families of gene segments called variable (V), diversity (D) and joining (J) are randomly assembled (so-called V(D)J recombination) into antibody variable-region genes during B-cell development² (Fig. 1). However, it has since become apparent that B cells can change their receptor specificity during development. Now, building on earlier evidence^{3–7}, Hertz *et al.*⁸ (reporting in *Nature* last month) and Meffre *et al.*⁹ (in the 17 August issue of the *Journal of Experimental Medicine*) suggest that mature B cells diversify their receptor repertoire in the antibody response. The cells do this by activating V(D)J recombination along with the well-established mechanism of somatic hyperpointmutation.

Somatic hyperpointmutation of antibody variable-region genes was the first mechanism of receptor modification to be discovered. It operates in T-cell-dependent antibody responses during proliferative expansion of the responding B cells in a particular microenvironment, the germinal centre. From the newly arising antibody mutants, those that show a high affinity for the immunizing antigen are selected into the 'memory' B-cell pool¹⁰.

A second process by which B cells can change receptor specificity is based on 'secondary' V(D)J rearrangements. Each newborn B cell expresses a distinct antigen receptor, with a distinct combination of a V_H, a D_H and a J_H gene segment for the variable region of the antibody heavy chain, and a V_L and a J_L segment for that of the light chain (V_k and J_k or V_λ and J_λ, depending on whether the cell expresses a light chain of class κ or λ)². However, any B cell that expresses a particular κ chain can continue, or reinitiate, rearrangements of the variable-region genes to get rid of the V_k/J_k gene segment initially expressed. It can then express a new κ chain, or switch to expression

of a λ chain. This is achieved either by rearrangements of upstream (non-rearranged) V_k genes to downstream (non-rearranged) J_k segments (Fig. 1), or by elimination of the V_k/J_k rearrangement carried by the cell, together with the κ constant-region gene on the same chromosome. Even in the case of the heavy chain, B cells can change specificity by rearranging a (non-rearranged) upstream V_H gene into the V_HD_H/J_H complex that was initially assembled and expressed¹⁰.

This change of receptor specificity by secondary variable-region gene rearrangements was initially identified as a process that eliminates autoreactive receptors in newly generated B cells in the bone marrow, and was termed receptor editing^{11,12}. Receptor editing seems to be triggered when the receptor is engaged by (self) antigen, and it is now considered to be an important mechanism of immunological tolerance and also antibody diversification. Only cells that fail to revise an autoreactive specificity are 'clonally' eliminated, as was postulated by the clonal-selection theory. But it was subsequently discovered that variable-region gene rearrangements can also be reinitiated in peripheral B cells — both *in vitro* through appropriate stimuli, and *in vivo*, in the germinal-centre reaction^{3–7}.

Hertz *et al.*⁸ and Meffre *et al.*⁹ now point to an important difference in the control of receptor editing in early B-cell development, compared with what happens in peripheral B cells and the germinal centre. In peripheral cells undergoing editing, when ligands bind the receptor with high avidity, variable-region gene rearrangements are terminated, not induced. Meffre *et al.* showed this for human tonsillar B cells activated by CD40 ligand and interleukins. Hertz *et al.* did similar experiments with B cells isolated from mice with transgenic receptors. When these mice were immunized with antigen that bound the transgenic receptor with either low or high avidity, the authors found that V(D)J recombination was induced only when the avidity was low. So, the purpose of secondary V(D)J recombination in T-cell-dependent antibody responses does not seem to be (as one might have expected) the elimination of potentially harmful self-

reactive antibody specificities arising in the germinal-centre reaction. Rather, V(D)J recombination diversifies the receptor repertoire in the responding cell population from which cells that bind antigen with high affinity are selected into the memory B-cell pool — similar to what is achieved by somatic hyperpointmutation.

Meffre *et al.* clearly show that, as in the mouse, gene rearrangements in human germinal-centre B cells are largely restricted to the 'centrocyte' subset, believed to originate from the proliferating 'centroblasts' in which somatic hyperpointmutation is occurring. But Hertz *et al.* document the initiation of V(D)J recombination in supposedly mature B cells *in vivo*, shortly (four days) after immunization, that is, before germinal centres usually develop. Thus, receptor diversification by secondary V(D)J recombination may be initiated before somatic hyperpointmutation.

We know that somatic hyperpointmutation occurs in most of the responding cells, but it is not clear to what extent receptor editing shapes the receptor repertoire upon antigenic stimulation. It could be restricted, for example, to a population of immature B cells drawn into the germinal-centre reaction — the editing cells resemble such cells to a surprising extent in terms of gene expression^{5,9}. However, although the generation of a new antibody repertoire by rearrangements of variable-region genes in response to antigen was initially a displeasing thought to many of us, it could help to recruit useful antibody specificities into the memory B-cell pool. Most of the diversity of pre-germinal-centre antibodies resides in the complementarity-determining region III of the heavy chain (encoded by a D element, adjacent non-templated nucleotides and parts of V and J), and this region may be the main determinant of antigen binding specificity¹³. So, receptor editing — which seems mainly to affect the κ-chain loci — could produce an (initial?) set of antibody variants that express distinct light chains, but still bind the nominal, or a related, antigen. This may be an optimal substrate for somatic hyperpointmutation to produce a range of high-affinity mutants in response to an invading pathogen. Indeed, in the chicken, the germinal-centre reaction seems to involve a similar interplay of somatic hyperpointmutation and gene conversion¹⁴.

Do these new results (and the process of somatic hyperpointmutation) challenge the clonal-selection theory, with its principle that antigen selects cells precommitted to an antibody specificity? Burnet, in his unique way, came close to where we now stand but did not seem to be happy with it¹: "The only question that really arises is whether, on a background of ... [clonal selection] ..., a new mechanism has been

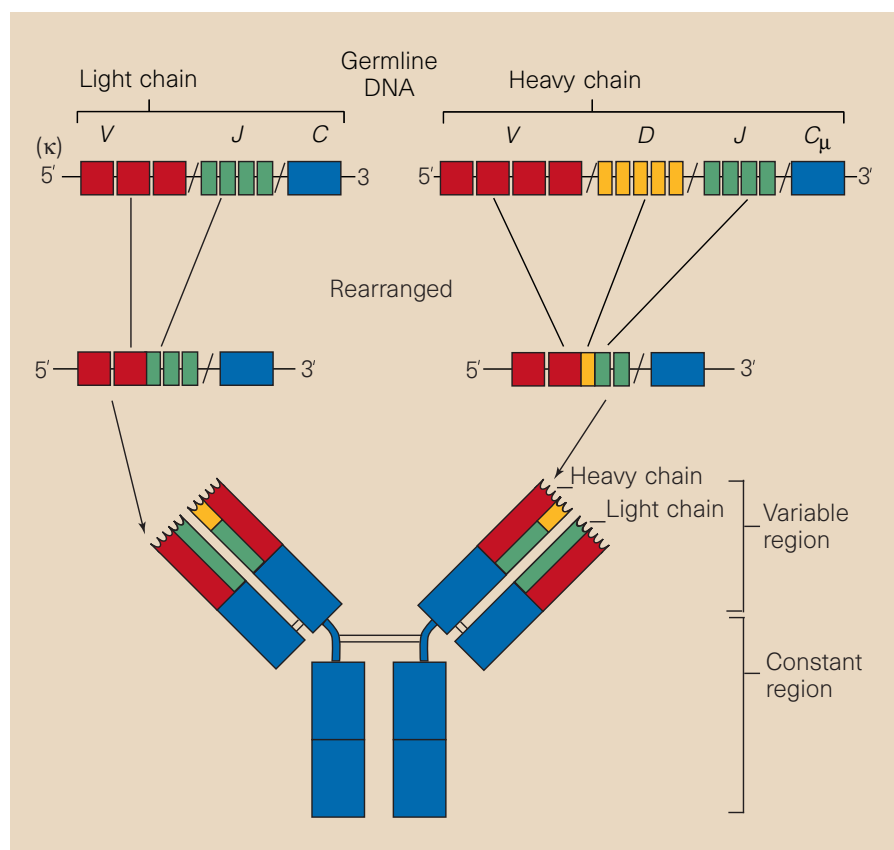


Figure 1 $V(D)J$ recombination in B-cell development. Antibodies comprise light and heavy chains, each of which is further subdivided into constant and variable regions. Rearrangement of the genes that make up the variable region allows the many different antibody specificities to be generated. Germline DNA contains many different variable (V), diversity (D) (in the case of the heavy chain only) and joining (J) gene segments (not all segments are shown). Secondary rearrangements can involve upstream, unrearranged V_{κ} , and downstream unrearranged J_{κ} elements, but other such rearrangements are possible.

superimposed by which production of an urgently needed protective substance may be accelerated. This might take the form of an enforcement of somatic mutation in the required direction in those cells whose natural pattern fitted best for this function. This would be a rather unhappy hybrid."

Present misunderstandings arise from what is hidden in the term 'in the required direction'. Burnet had no objection to the cells changing their receptor specificity by somatic mutation during their selection by antigen; in fact, this was an important element of his theory. But 'in the required direction' relates to the 'template' function of antigen postulated at the time by the instructionalists — the concept that Burnet tried to demolish. He would have liked the idea that the cells responding to antigen had found efficient genetic means to diversify their receptor repertoire, such that the occurrence of high-affinity mutants in the population would be increased to improve clonal selection. And, in the context of receptor editing, he would have asked how the pool of V_L genes could have been selected in evolution so that light-chain editing

in the antibody response would allow a more efficient defence against pathogens. Thus, Burnet's 'unhappy hybrid' is not unhappy as it turns out, in the sense that it is not truly a hybrid between two opposing theories. □

Klaus Rajewsky is at the Institute for Genetics, University of Köln, Weyertal 121, D-50931 Köln, Germany, and the EMBL Mouse Biology Programme, Via Ramarini 32, I-00016 Monterotondo-Scalo (Rome), Italy.
e-mail: klaus.rajewsky@genetik.uni-koeln.de

1. Burnet, F. M. *The Clonal Selection Theory of Acquired Immunity* (University Press, Cambridge, 1959).
2. Tonegawa, S. *Nature* **302**, 575–581 (1983).
3. Han, S. *et al. Science* **274**, 2094–2097 (1996).
4. Hikada, M. *et al. Science* **274**, 2092–2094 (1996).
5. Han, S. *et al. Science* **278**, 301–305 (1997).
6. Papavasiliou, F. *et al. Science* **278**, 298–301 (1997).
7. Hikida, M. & Ohmori, H. *J. Exp. Med.* **187**, 795–799 (1998).
8. Hertz, M., Kouskoff, V., Nakamura, T. & Nemazee, D. *Nature* **394**, 292–295 (1998).
9. Meffre, E. *et al. J. Exp. Med.* (in the press).
10. Rajewsky, K. *Nature* **381**, 751–758 (1996).
11. Tiegs, S. L. *et al. J. Exp. Med.* **177**, 1009–1020 (1993).
12. Gay, D. T. *et al. J. Exp. Med.* **177**, 999–1008 (1993).
13. Davis, M. M. *et al. Annu. Rev. Immunol.* **16**, 523–544 (1998).
14. Arakawa, H., Furusawa, S., Ekino, S. & Yamagishi, H. *EMBO J.* **15**, 2540–2546 (1996).

Daedalus

Bubble power

The internal-combustion engine is a very imperfect device. It breathes air, burns fuel in it imperfectly, and returns highly polluting gases to the atmosphere. Daedalus sees this gas-phase reaction as the central problem. He is trying a new twist: he wants the engine to burn an aqueous foam.

A foam of inflammable gas mixture has many advantages. First, the engine could be fed with several streams of foam, each with its own concentration of fuel. The stratified charge, that dream of motor engineers, would be a reality at last. The centre of each cylinder could receive a rich mixture, while the walls received a much more dilute one, or even pure air. Flame would then never touch the walls. Pollutants, mainly created by premature quenching of the reaction on the cylinder surfaces, would never form.

When a combustible foam burns, its bubbles are shocked into bursting before the flame reaches them. What actually burns is a gas mixture loaded with aqueous droplets. These should go to steam, which would usefully increase the pressure of the expanding gas while moderating its temperature. Water injection can be used to increase the thrust of jet engines at take-off, and has often been advocated for internal-combustion engines.

The exhaust stream, of course, would be a pure hot gas. But Daedalus would like to enclose that in foam as well. Particulates and acidic combustion products would be trapped in the foam walls. They would be discharged not into the air, but as a ribbon of foam on the road, ultimately to be washed into the drainage system. It would provide a temporary white road marking, indicating to following traffic the presence, track and probable distance of the foam-laying vehicle ahead — a useful adjunct to road safety.

To complete the scheme, Daedalus would like to lubricate the engine with the same aqueous detergent system from which the foam is generated. This is a tough technical challenge. With good fortune, a foam engine would run so much cooler than a conventional one that the sliding surfaces could be made not of metals (for which hydrocarbon oils seem the only reliable lubricants) but from specialized plastics or ceramics. Ignition is an even tougher challenge; no sparking plug will work in the wet. A chemical system, based perhaps on spontaneous igniters such as silicon or phosphorus hydrides, may be needed.

David Jones

Obituary

Colin Patterson (1933–98)

Palaeontologist – reformer of the fossil record

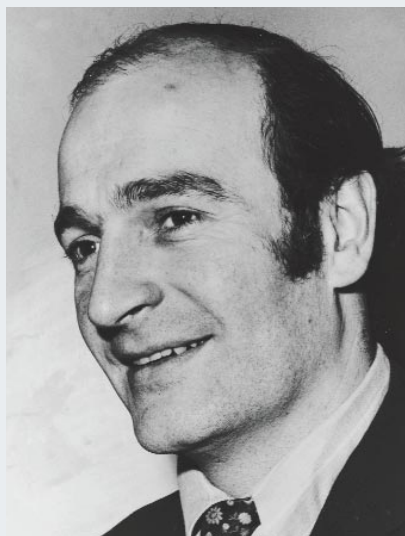
Colin Patterson, who died in London earlier this year at the age of 64, will be remembered for his part in the cladistic reform of palaeontology. This was the period in which the traditional method in palaeontology, the search for ancestors, was abandoned in favour of the search for the sister group — of evidence of the nearest relative.

Patterson was born in Hammersmith, west London, in 1933, and educated at Tonbridge school and Imperial College, London. His first appointment was as a lecturer at Guy's Hospital Medical School. While doing that job, he finished his PhD thesis on Mesozoic teleost fishes, which was published by the Royal Society in 1964.

It was this and other early studies of fossil fishes and their anatomy that led to Patterson's decisive response to the challenge of cladistics as it developed in the late 1960s. His response took years to mature and it was achieved through collaboration with colleagues working in museums in Europe and the United States. Eventually, he came to that vantage point of understanding described by Thomas Paine (1782): "We see with other eyes; we hear with other ears; and we think with other thoughts, than those we formerly used. We can look back on our own prejudices, as if they had been the prejudices of other people."

The challenge came in 1966 in a monograph by the senior entomologist of Naturhistoriska Riksmuseet in Stockholm, Lars Brundin: "Transantarctic relationships and their significance, as evidenced by chironomid midges", published by the Royal Swedish Academy of Sciences. Brundin summarized phylogenetic principles as developed by Willi Hennig, and critically appraised world biogeography as summarized by Philip Darlington.

Brundin's work became a favoured topic during informal and peripatetic debate at the Natural History Museum in South Kensington — then the British Museum (Natural History), where, in 1962, Patterson had been appointed scientific officer in the department of palaeontology. Discussions took place once or twice daily under the museum's colonnade, the backdoor service area where open flame and resulting tobacco smoke were tolerated. Patterson later wrote of those times that "After 10 years' work in that field, I read Brundin, and still recall the



excitement with which I realized that there is a logical basis to evolutionary relationships which I had never seen discussed".

Such evolutionary relationships include those between ancestors and descendants. Patterson explained that it is this type of connection that fossils are expected to document. When viewed as the relationship between two groups, descent means that one (the ancestral group) is paraphyletic — characterized only by lack of homologies rather than their presence. He soon came to see all alleged ancestral groups as "paraphyletic, and therefore unreal, obscuring rather than solving questions of evolutionary relationship.... Why basal ancestral groups should be an effective bar to progress finally became clear [in 1966] with the publication ... of Hennig's *Phylogenetic Systematics*, and of Brundin's exposition of Hennig's methods."

Patterson credited Hennig with the discovery of paraphyly, but it was Patterson, as much as or more than anyone else, who realized the implications of the equation of ancestry and paraphyly for the fossil record generally. This he accomplished through his empirical work, and in review articles on ichthyology, palaeontology and biogeography, and on the contrasts between morphological and molecular biology. In detail of description and eloquence of interpretation and argument, particularly in oral presentation, he was without peer. He asked, for example, "Is it not strange that the justification of phylogeny, as something beyond systematics, resides in extinct paraphyletic groups? For those groups are the inventions of evolutionists, those who appeal to them as demonstrating the path of descent." With Patterson in mind, Brundin commented

that, "Little by little some palaeontologists have perceived that Hennig's principles of phylogenetic systematics meant a revolution to their science".

Revolution provokes counter-revolution. Readers may remember an appalled Beverly Halstead and his thundering commentaries in the late 1970s and early 1980s, which with editorial blessing later became directed towards cladistics as reflected in Halstead's "Museum of errors" (*Nature* 288, 208; 1980). The target was the exhibits on dinosaurs and humans, then on display at the Natural History Museum, with explanatory texts providing interpretations according to cladistic principles. For Halstead these "present[ed] the public for the first time with the notion that there are no actual fossils directly antecedent to man. What the creationists have insisted on for years is now being openly advertised by the Natural History Museum." Creationists took notice, but the sky neither trembled nor fell — except upon Little Essex Street, then the site of *Nature's* editorial office, in the form of vigorous responses from Patterson and many others.

In the mid-1980s Patterson avidly embraced the possibilities offered by molecular systematics. Too hopefully, it seems in retrospect, he saw in the molecularly based neutral theory of evolution an equivalence in the methods of phenetics and cladistics: "It should follow that the prospects for a unified discipline of systematics are excellent; clocks and clades show the way forward". Ten years later, he ended a review of molecules versus morphology (with two colleagues of the museum): "As morphologists with high hopes of molecular systematics, we end this survey with our hopes dampened. Congruence between molecular phylogenies is as elusive as it is in morphology, and as it is between molecules and morphology."

Although he formally retired from the Natural History Museum in October 1993, Patterson continued working there as an honorary research fellow. On 9 March, however, he was mortally stricken with heart failure. His professional life is summarized, and his many publications are listed, in the first chapter of *Interrelationships of Fishes*, published by Academic Press in 1996. That legacy will last for many years to come.

Gareth Nelson

Gareth Nelson, formerly of the American Museum of Natural History, is in the School of Botany, University of Melbourne, Parkville, Victoria 3052, Australia.
e-mail: nelson@amnh.org

Attractive attractors

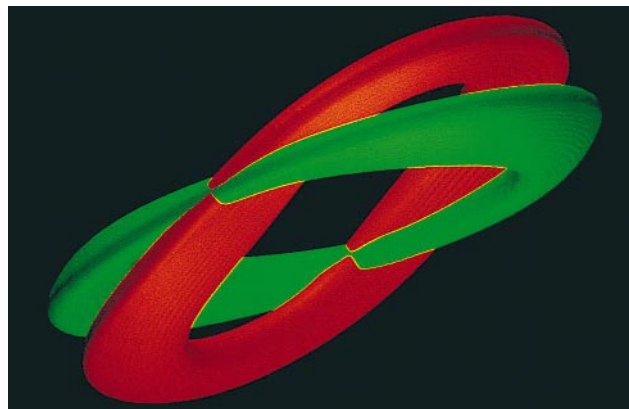
It can be easier to understand mathematical procedures when we can see them in, seemingly, three dimensions. Even though we cannot really see all the dimensions in which these attractors exist, they are satisfying to the eye.

Martin Kemp

For those of us who can best cope (or can only cope adequately) when mathematical procedures can be envisaged as processes in space, the graphic plotting of chaotic systems comes as something of a relief, reclaiming visual territory lost to the logicians of symbolic computation since the era of Descartes. The algebra of chaos bears fruit graphically in the illusory space behind the computer screen, yielding the strangely beautiful secrets of its attractors and fractal self-similarities to new virtuosi of the electronic palette.

Yet there is a paradox here. The spatial and coloristic magic of the graphics of chaos, to which we respond with senses attuned to plastic forms in our own experimental world are products of a quite different realm of conceptual space in which the coordinates of normal three-dimensionality stand incompletely and symbolically for multi-dimensional coordinates of which the three standard directions of space provide only one set.

A conscious search for an effective mode for the rendering of a chaotic system was



Attractors for coupled pendulum in regular motion (Tom Mullin and Anne Skeldon).

undertaken in 1990 by Anne Skeldon and Tom Mullin of the Non-Linear Systems Group, then at the Clarendon Laboratory in Oxford. They chose the coupled pendulum, a now classic exemplar of how simple deterministic means can, under particular conditions, result in surprising unpredictability.

From the end of a pendulum that can swing in one plane is hung a second pendulum that swings at right angles to it. The pivot of the upper pendulum is driven up and

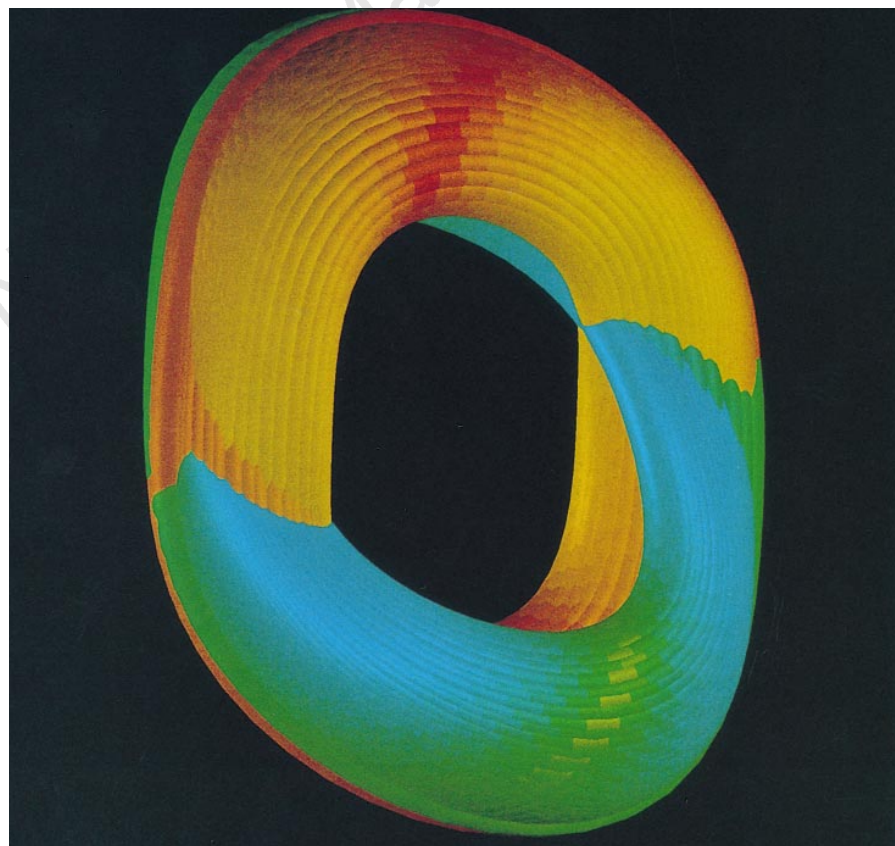
down at a regular frequency. Five coordinates are chosen for plotting: the velocity of the top pendulum, its angle, the velocity of the drive, the velocity of the lower pendulum, and its angle. As the frequency of the drive changes, complex, chaotic assemblies of periodic motions are generated. Three of the five coordinates are selected: the angle of one pendulum; the angle of the other and the amplitude of the drive. Plotting the trajectories in the phase space as a three-dimensional graph, two characteristic attractors emerge, each in the now familiar form of a torus. Within the conventions of the mapping, which displays the torus bands as intersecting and assigns different colours to differentiate them, they merge progressively at lower frequencies.

It is possible to vary the colour within the torus bands to map a fourth dimension — in this instance speed of motion — so that an additional coordinate for the behaviour of the system can be graphically represented. The colours are in a sense arbitrary, yet the encoding red as fast and blue as slow is no less contrived than our automatic registering of frequencies of light as different colours.

The attractors exhibit undeniable appeal as plastic artefacts, yet the true space they occupy can only be seen by what Feynman called “the eye of analysis”, rather than the sensory eye.

Evolution has equipped us with sensory eyes for the mastery of phenomenological space. It is these material eyes' enduring demands for gratifying comprehension that are powerfully satisfied by the translation of multi-dimensional variables into a visual realm that our limited apparatus can knowingly see. □

*Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.
e-mail: martin.kemp@trinity.oxford.ac.uk*



Attractors for coupled pendulum in chaotic motion. The dimension of speed is given by red = fast and blue = slow (Tom Mullin and Anne Skeldon).

Cretaceous plesiosaurs ate ammonites

The Plesiosauria, large Mesozoic marine reptiles, have been regarded as dominant predators in Mesozoic seas, although there is only a little (including some doubtful) direct evidence of their stomach contents^{1–3}. Here we report the occurrence in Hokkaido, northern Japan, of a partial skeleton of a Cretaceous short-necked plesiosaur together with a large number of isolated ammonoid jaws. This is, to our knowledge, the first description of the stomach contents of a polycotyloid plesiosaur, and the first firm evidence of prey–predator relationships between plesiosaurs and ammonites.

We excavated the partial skeleton (specimen UMUT MV 19965, University Museum, University of Tokyo, Tokyo, Japan) from an Upper Cenomanian (Upper Cretaceous) outcrop along the Obirashibe River in Hokkaido⁴. The skeleton consists of the right half of the pectoral girdle, almost complete right forelimb, disarticulated vertebrae, several gastralia, and fragments of the pelvic girdle. Unusually for a short-necked plesiosaur, gastroliths were preserved in the stomach region. Associated fossils include about 30 isolated and disarticulated cephalopod jaw apparatus, a shark's tooth, and molluscan shells and moulds.

The way in which the jaw apparatus occur indicates that they were contents of the plesiosaur gut⁵. All of the isolated cephalopod jaws, together with gastroliths, are concentrated in the stomach region of the plesiosaur. In the Upper Cretaceous stages of Hokkaido, isolated cephalopod jaws are occasionally found together with shells of ammonites and inoceramid bivalves in calcareous concretions, but they rarely occur as clusters as they do here. In this case, the fossil-bearing concretion (Fig. 1a) and surrounding mudstone lack any distinct sedimentary structures. These lines of evidence strongly suggest that the cephalopod jaws found with the plesiosaur specimen did not accumulate accidentally by means of currents, but were preserved as dietary remains within the plesiosaur digestive tract.

All the cephalopod jaws studied are small, measuring 5–15 mm in maximum length (Fig. 1). They are made of a black horny material and lack any trace of calcification. Following the criteria by which to distinguish between upper and lower jaws of cephalopods^{6–8}, we identified four (three lower and one upper) relatively well-preserved jaws.

The outer lamella in two of the three lower jaws is ornamented with prominent radial striations (Fig. 1d–f), whereas that in the third lacks apparent surface architecture (Fig. 1g), indicating that two different taxa may be present. The lower jaws of

cephalopods can be classified into two morphotypes: one with a gently opening larger outer lamella and a short, reduced inner lamella, and the other with a widely open, laterally projected outer lamella and a long and posteriorly projected inner lamella. The former type is known in extant *Nautilus* and ammonites, whereas the latter type occurs in coleoids. All three of the lower jaws studied are of the former type.

The preserved upper jaw consists of a pair of wide inner lamellae and a short, reduced outer lamella, which are combined in the sharply pointed anterior portion (Fig. 1b, c). The paired inner lamellae are ornamented with many radial striations. Among extant and fossil cephalopods, upper jaws with paired inner lamellae are

known only from Jurassic and Cretaceous ammonites, excluding the *Lytocerotina*.

Comparison of the morphology of these jaws with the jaws of extant and fossil cephalopods^{6–8} shows that our specimens are referable to ammonites, and the size of the jaws indicates that the ammonite shells were probably relatively small⁸. The exact taxonomic relationships of these specimens are unknown, although the overall morphology of the jaws is similar to that of the jaws of desmoceratid ammonites⁹.

The reason why the ammonites lack calcareous shells remains unknown. The effects of stomach acid and/or preferential selection by the plesiosaur before swallowing may be responsible. The fact that no skull is present with the excavated plesio-

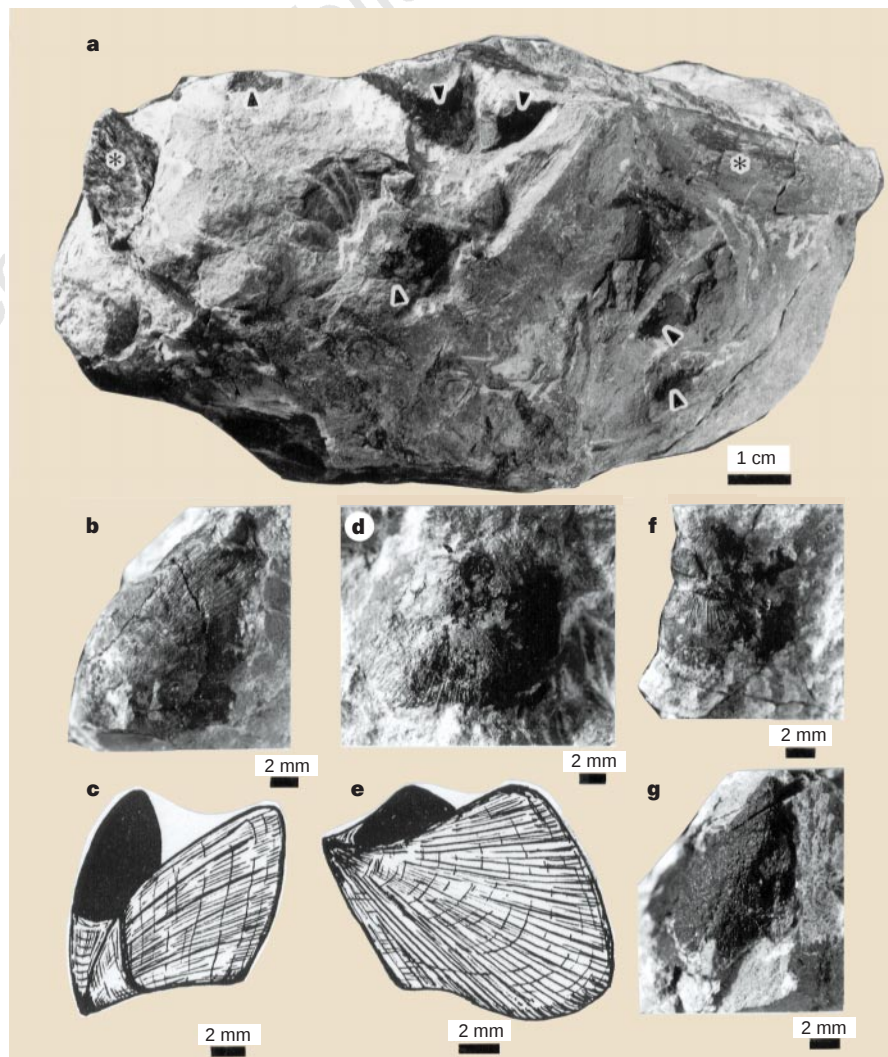


Figure 1 Cephalopod jaws referable to ammonoids found at the posterior end of the coracoid of specimen UMUT MV 19965. **a**, Photograph of the concretion slab, showing the distribution of cephalopod jaws (arrow heads) and plesiosaur gastralia (asterisks). **b**, Photograph of a cephalopod upper jaw, specimen UMUT MV 19965 (lateral view). **c**, Drawing of the jaw shown in **b** (anterolateral view). **d**, Photograph of a cephalopod lower jaw, specimen UMUT MM 19998 (frontal view). **e**, Drawing of the jaw shown in **d** (anterolateral view). **f**, Photograph of a cephalopod lower jaw, specimen UMUT MM 19999 (anterolateral view). **g**, Photograph of a cephalopod lower jaw, specimen UMUT MM 20000 (anterior view). The left half is missing.

saur makes it difficult to speculate about the functional morphology of the predator's teeth. However, the known specimens of Polycotylidae of comparable size (about 3 metres) have relatively slender teeth, which are poorly adapted to crush ammonite shells. The inferred small size of the present specimens of prey ammonites indicates that they might have been swallowed whole, even if the plesiosaur's teeth were not suitable for crushing shells.

This direct evidence of a plesiosaur's diet, together with corresponding studies of tooth morphology¹, offers an insight into the prey preferences of these marine reptiles.

Tamaki Sato*

Department of Geology, University of Cincinnati,
PO Box 210013, Cincinnati,
Ohio 45221-0013, USA

*Present address: Department of Geology and
Geophysics, University of Calgary,
2500 University Drive NW, Calgary,
Alberta T2N 1N4, Canada
e-mail: BXF02315@nifty.ne.jp

Kazushige Tanabe

Geological Institute, Graduate School of Science,
University of Tokyo, Hongo 7-3-1,
Tokyo 113-0033, Japan

- Massare, J. A. *J. Vert. Paleontol.* **7**, 121–137 (1987).
- Brown, B. *Science* (N.S.), **20**, 184–185 (1904).
- Matsumoto, T., Obata, I., Okazaki, Y. & Kanie, Y. *Proc. Jpn. Acad. B* **58**, 109–113 (1982).
- Sato, T. *J. Vert. Paleontol. Abstr.* **17**, 73 (1997).
- Sato, T. *J. Vert. Paleontol. Abstr.* **17**, 73 (1997).
- Clarke, M. R. *A Handbook for the Identification of Cephalopod Beaks* (Clarendon, Oxford, 1986).
- Nixon, M. in *Ammonoid Paleobiology* (eds Landman, N., Tanabe, K. & Davis) 23–42 (Plenum, New York, 1996).
- Tanabe, K. & Fukuda, Y. in *Functional Morphology of the Invertebrate Skeleton* (ed. Savazzi, E.) (John Wiley, in the press).
- Tanabe, K. *Palaeontology* **26**, 677–686 (1983).

Connexin mutations in deafness

Genetic deafness is one of the most prevalent inherited sensory disorders, affecting about 1 in 2,000 children. Mutations in the connexin 26 gene have been associated with autosomal recessive non-syndromic deafness (DFNB1)¹. The connexin 26 gene is a member of the connexin family of genes, which encode intercellular channels comprising gap junctions², and it is abundantly expressed in the organ of Corti^{1,3}. Here we test the channel-forming ability of mutant connexin 26 proteins using a well-characterized *in vitro* system for functional expression of connexin channels⁴. We find that mutant connexin 26 proteins can act as dominant inhibitors of wild-type connexin 26 channel activity.

After the first identification of mutations in the connexin 26 (*Cx26*, or *GJB2*) gene in association with autosomal recessive non-syndromic deafness¹, several more groups

Table 1 Conductance of wild-type and mutant Cx26 channels in paired *Xenopus* oocytes

Cell injection	Mean conductance (pS)	Standard error	Number of pairs
Cell 1/cell 2			
H ₂ O/H ₂ O	0.007*	0.002	12
Cx26/Cx26	1.57	0.50	23
M34T/M34T	0.006*	0.002	12
M34T + Cx26/M34T + Cx26	0.046*	0.019	11
W77R/W77R	0.005*	0.002	6
W77R + Cx26/W77R + Cx26	1.11	0.55	5

*Differences in conductance values were statistically significant ($P < 0.05$) compared with wild-type Cx26 data (Student's *t*-test).

showed that mutations in this gene cause recessive (and sporadic) non-syndromic deafness^{3–9}. A *Cx26* mutation has also been associated with a dominant form of non-syndromic deafness (DFNA3) in a single pedigree¹. However, this finding has been challenged by Scott *et al.*¹⁰ who identified a family in which heterozygote carriers of that allele do not exhibit hearing loss.

We studied the controversial dominant *Cx26* mutant M34T, in which a methionine is converted to threonine^{1,10}, and a well-characterized recessive mutant, W77R, in which a tryptophan is converted to arginine⁵. These mutants are similar in that they encode full-length products containing non-conservative amino-acid substitutions.

Xenopus laevis oocytes were injected with *in vitro*-transcribed complementary RNA encoding wild-type human *Cx26*, mutant *Cx26*, or combinations of wild-type and mutant *Cx26* proteins. After the removal of the vitelline envelope, oocytes were paired, and intercellular conductance was measured directly by dual voltage clamp 48 hours later (Table 1).

Expression of wild-type *Cx26* resulted in robust conductances, indicating high levels of intercellular channel activity. In contrast, the M34T variant did not induce coupling between paired oocytes above background levels (measured in water-injected control pairs). More important, expression of the M34T mutant with wild-type *Cx26* resulted in marked inhibition ($>95\%$, $P < 0.05$) of intercellular coupling. The recessive W77R

mutant also failed to induce intercellular channel activity, but did not significantly inhibit ($\sim 29\%$, $P > 0.05$) the ability of wild-type *Cx26* to form functional channels when coexpressed.

Following electrophysiological measurements, oocyte extracts were western-blotted to confirm that mutant and wild-type forms of the *Cx26* protein were produced at similar levels (Fig. 1).

These data support the hypothesis that the M34T mutation causes hearing loss by dominant inhibition of the activity of wild-type *Cx26* channels. In addition, the loss of function without dominant inhibition exhibited by the W77R mutant concurs with the recessive inheritance of this allele. However, the normal hearing in the M34T carriers from the pedigree described by Scott *et al.*¹⁰ cannot yet be explained. Although Scott *et al.* thought it unlikely, mechanisms such as second-site mutations that silence the allele or compensatory changes in the expression of other genes might explain the difference in hearing status between the M34T gene carriers from the two pedigrees^{1,10}. Analysis of M34T allelic expression in the pedigree with normal hearing may help to resolve this issue.

A different *Cx26* mutant, in which tryptophan 44 is altered to cysteine, has recently been associated with autosomal dominant deafness in multiple families¹¹, providing compelling support for the identification of *Cx26* as the gene altered in both DFNB1 and DFNA3. Moreover, our results provide a molecular explanation for the dominant effect of the M34T allele.

Thomas W. White, Michael R. Deans

Department of Cell Biology,
Harvard Medical School,
Boston, Massachusetts 02115, USA

David P. Kelsell

Centre for Cutaneous Research,
St Bartholomew's and the Royal London Hospital,
London E1 2AT, UK

David L. Paul

Department of Neurobiology,
Harvard Medical School,
Boston, Massachusetts 02115, USA
e-mail: dpaul@hms.med.harvard.edu

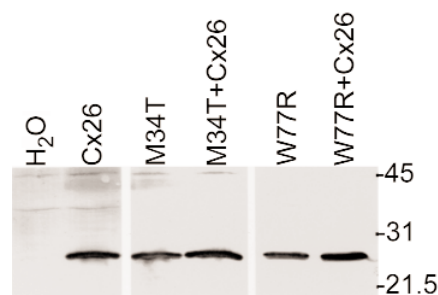


Figure 1 Western blots of *Xenopus* oocyte extracts indicate that wild-type and mutant connexins are expressed at similar levels. Ten oocytes from each injected group were pooled, and membrane fractions were prepared¹². Following resuspension in sample buffer, 20% of the preparation was loaded in each lane for each condition. The western blot was probed with an affinity-purified rabbit anti-Cx26 antibody¹³.

- Kelsell, D. P. *et al.* *Nature* **387**, 80–83 (1997).
- Bruzzzone, R., White, T. W. & Paul, D. L. *Eur. J. Biochem.* **238**, 1–27 (1996).
- Kikuchi, T., Kimura, R.S., Paul, D.L. & Adams, J. C. *Anat. Embryol.* **191**, 101–118 (1995).

4. Swenson, K.L., Jordan, J. R., Beyer, E. C. & Paul, D. L. *Cell* **57**, 145–155 (1989).
5. Carrasquillo, M. M., Zlotogora, J., Barges, S. & Chakravarti, A. *Hum. Mol. Genet.* **6**, 2163–2172 (1997).
6. Zelante, L. et al. *Hum. Mol. Genet.* **6**, 1605–1609 (1997).
7. Denoyelle, F. et al. *Hum. Mol. Genet.* **6**, 2173–2177 (1997).
8. Estivill, X. et al. *Lancet* **351**, 394–398 (1998).
9. Lench, N., Houseman, M., Newton, V., Van Camp, G. & Mueller, R. *Lancet* **351**, 415 (1998).
10. Scott, D. A., Kraft, M. L., Stone, E. M., Sheffield, V. C. & Smith, R. J. *Nature* **391**, 32 (1998).
11. Denoyelle, F. et al. *Nature* **393**, 319–320 (1998).
12. White, T. W., Bruzzone, R., Goodenough, D. A. & Paul, D. L. *Mol. Biol. Cell* **3**, 711–720 (1992).
13. Goliger, J. A. & Paul, D. L. *Dev. Dyn.* **200**, 1–13 (1994).

Very long carbon nanotubes

Carbon nanotubes¹ can now be produced in large quantities by either arc methods^{2,3} or thermal decomposition of hydrocarbons^{4,5}. Here we report that pyrolysis of acetylene over iron/silica substrates is an effective method with which to produce very long, multiwalled carbon nanotubes that reach about 2 mm in length, which is an order of magnitude longer than that described in most previous reports^{1–5}.

The method we used is similar to that reported in ref. 5, but we have made some improvements to substrate preparation. We prepared the substrates by a sol–gel process using the following technique. We mixed tetraethoxysilane (10 ml) with 1.5 M iron nitrate aqueous solution (15 ml) and ethyl

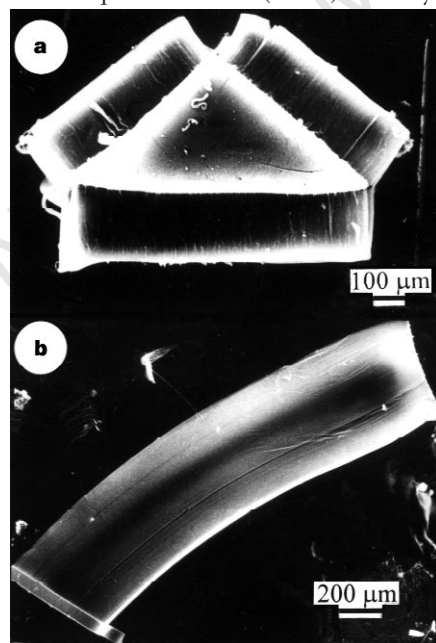


Figure 1 Low-magnification scanning electron microscope images of aligned carbon nanotube arrays. **a**, Top view of a sample after 5 hours of growth. The triangular substrate is covered with nanotube arrays growing outwards perpendicularly from every surface of the substrate. **b**, A sample after 48 hours of growth. Only one nanotube array remains after stripping off other arrays from the substrate.

alcohol (10 ml) by magnetic stirring for 20 minutes. We added a few drops of concentrated hydrogen fluoride (0.4 ml), and the mixture was stirred for another 20 minutes. The mixture was then dropped onto a quartz plate to form a film 30–50 μm thick. After gelation, the film was dried overnight at 80 °C, during which time the gel cracked into small pieces of substrate of area 5–20 mm². The substrates were calcined at 450 °C for 10 hours under vacuum and then reduced at 500 °C for 5 hours in a flow of 9% H₂/N₂ under 180 torr.

At this stage, large quantities of iron/silica nanoparticles, which acted as catalysts for nanotube growth, formed evenly on all surfaces of the substrates. Subsequently, carbon nanotubes were produced on the substrates in a flow of 9% acetylene in nitrogen at 600 °C for 1–48 hours under 180 torr.

We used this method to prepare carbon nanotubes at a very high yield. Every surface of the substrate was covered with a nanotube array composed of large quantities of highly aligned nanotubes (Fig. 1a). The lengths of the nanotube arrays increased with growth time, and reached about 2 mm (Fig. 1b) after 48 hours of growth. We believe the nanotubes could be even longer if the growth time was further increased.

High-magnification scanning electron microscope images (Fig. 2a) show that the carbon nanotubes grow outwards separately and perpendicularly from the substrate to form an array. The nanotubes within these arrays are of uniform external diameter (~20–40 nm), with a spacing of about 100 nm between the tubes. Most of the nanotubes in the arrays are highly aligned, although a few appear to be slightly tangled or curved.

The nanotube arrays can be easily stripped off from the substrates, and scanning electron microscope observations of the surfaces of the resulting substrates show that the iron/silica nanoparticles that were present on the surfaces of the substrates before nanotube growth disappear after such growth. But subsequent reduction of the substrates after nanotube growth resulted in the transportation of more catalyst particles to the substrate surfaces from within the bulk substrate. Therefore, the substrates can be re-used after reduction to grow new nanotube arrays with characteristics similar to the original ones.

High-resolution transmission electron microscope observations (Fig. 2b) show that the multilayered graphite tubes (10–30 layers) are coated with a small amount of amorphous carbon on their periphery owing to the low growth temperature (600 °C) and long growth time.

To determine whether the long nanotubes grow continuously from the bottom

to the top within the nanotube arrays, we cleaved some thinner nanotube bundles (Fig. 2c) from the arrays and measured their resistivity (ρ) at temperatures from room temperature to 10 K by a four-probe method. The temperature dependence was semiconductor-like for most samples, but metallic behaviour was also seen ($\rho \approx 10^{-2}$ to 10^{-3} Ωcm and $|\Omega d\rho/dT\Omega| \approx 10^{-5}$ to 10^{-6} Ωcm K⁻¹ at room temperature). This resistivity is an order of magnitude larger than the values reported previously^{6,7}, presumably because of the presence of large numbers of defects in our very long nanotubes. These results indicate that these very long nanotubes do indeed grow continuously without any interruption within the nanotube arrays.

Possible uses of the nanotubes are related mainly to their dimensions and electronic⁸ and mechanical^{9–11} properties. The main factors that may limit their practical use are the cost, yield and quality. Very long nanotubes may be useful in, for example, the probe tips of scanning tunnelling microscopes¹², field emission materials¹³ and nanotube-reinforced materials^{9–11}. We are now

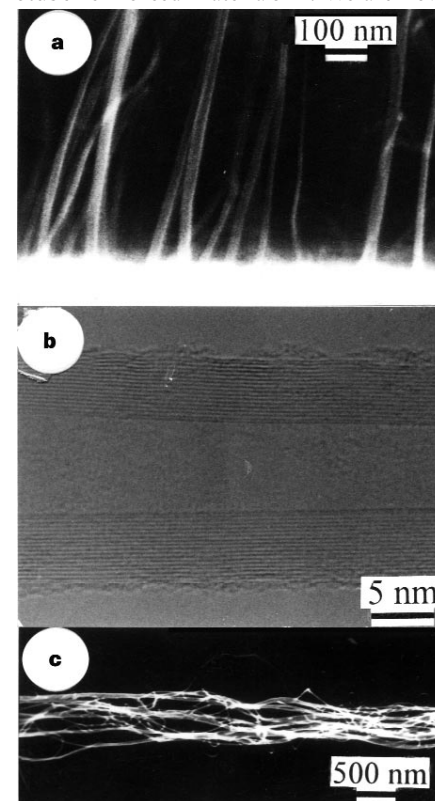


Figure 2 Scanning and transmission electron microscope images of nanotubes after 48 hours of growth. **a**, High-magnification scanning electron microscope image of an aligned carbon nanotube array. **b**, High-resolution transmission electron microscope image of a carbon nanotube. **c**, Scanning electron microscope image of a thin nanotube bundle cleaved from a nanotube array along the growth direction of the nanotubes. The bundle contains about 20 nanotubes. The total length of the

attempting to measure the tensile strength of very long nanotubes.

Z. W. Pan, S. S. Xie*, B. H. Chang,
C. Y. Wang, L. Lu, W. Liu,
W. Y. Zhou, W. Z. Li

Institute of Physics, Chinese Academy of Sciences,
Beijing 100080, China

* e-mail: user412@aphy01.iphy.ac.cn

L. X. Qian

Department of Physics,
Central University of Nationalities,
Beijing 100081, China

1. Iijima, S. *Nature* **354**, 56–58 (1991).
2. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
3. Bethune, D. S. *et al.* *Nature* **363**, 605–607 (1993).
4. Endo, M. *et al.* *Carbon* **33**, 873–881 (1995).
5. Li, W. Z. *et al.* *Science* **274**, 1701–1703 (1996).
6. Ebbesen, T. W. *et al.* *Nature* **382**, 54–56 (1996).
7. Fisher, J. E. *et al.* *Phys. Rev. B* **55**, 4921–4924 (1997).
8. Dai, H. J., Wong, E. W. & Lieber, C. M. *Science* **272**, 523–526 (1996).
9. Treacy, M. M. J., Ebbesen, T. W. & Gibson, J. M. *Nature* **381**, 678–680 (1996).
10. Wong, E. W., Sheehan, P. E. & Lieber, C. M. *Science* **277**, 1971–1975 (1997).
11. Falvo, M. R. *et al.* *Nature* **389**, 582–584 (1997).
12. Dai, H. J. *et al.* *Nature* **384**, 147–150 (1996).
13. de Heer, W. A., Chatelain, A. & Ugarte, D. *Science* **270**, 1179–1181 (1995).

Switch from specialized to generalized pollination

The once prevalent view that the evolution of extreme ecological specialization is accompanied by a loss of potential for adapting to new conditions, and thus is irreversible^{1–4}, has been challenged by several recent examples^{1,2,5}. However, we know of no modern phylogenetic studies showing reversal in pollination relationships from extreme specialization to generalization, although such reversals are theoretically expected^{6,7}. Here we present molecular phylogenetic evidence for an evolutionary shift in *Dalechampia* (Euphorbiaceae) vines from a highly specialized relationship (pollination by one or a few animal species^{2,7}) with resin-collecting bees to generalized pollination by a variety of pollen-feeding insects. This shift was associated with dispersal from Africa to Madagascar, where the specific resin-collecting pollinators are absent. These results show that plants dispersing beyond the range of their specific pollinators may succeed by evolving more generalized pollination systems.

Only a few genera of bees in the families Megachilidae and Apidae use floral resin in nest construction^{8,9}, and only two genera of plants, *Dalechampia* and *Clusia* (Clusiaceae), are known to secrete resins that attract pollinators⁸. Partly because the reward they offer is non-nutritive, *Dalechampia* blossoms that secrete resin are pollinated by only one or two species of bees at any one location, and so depend on

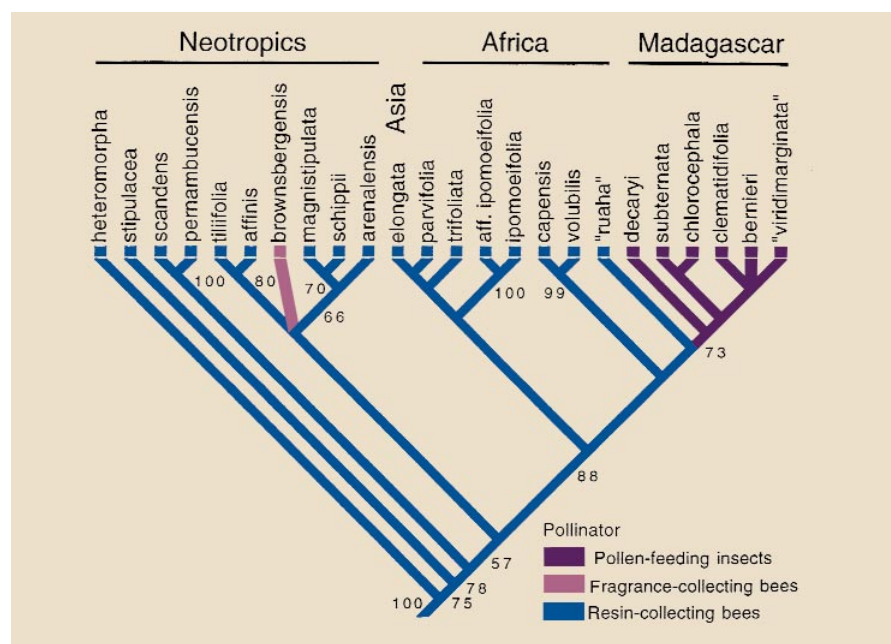


Figure 1 A phylogenetic hypothesis for *Dalechampia* sects. *Dalechampia* and *Tiliifoliae* based on maximum parsimony (MP) analysis of combined nuclear ribosomal (ITS-1, 5.8S, ITS-2) and chloroplast (*trnK* intron) DNA sequences. The strict consensus of MP trees is depicted, with bootstrap values (above 50%) along branches. The tree shown is part of a larger tree from an analysis including 16 representatives of the other sections of *Dalechampia* and two species from candidate sister genera (*Plukenetia* and *Tragia*). Evolution of pollination ecology was mapped using MP onto the tree based on the observed pollination of each species. The species are endemic to the regions indicated except *D. parvifolia*, which occurs in both Africa and Asia.

specific pollinators. In eastern and southern Africa, for example, *Dalechampia* populations are usually pollinated by only one species of *Pachyanthidium* or *Heriades* (Megachilidae)^{9,10}. Recent fieldwork in Madagascar has shown that the species of *Dalechampia* found there offer only pollen as a reward for pollinators, and that most are pollinated by a variety of pollen-feeding insects, including beetles (Cerambycidae, Scarabidae), muscoid flies (Diptera) and several bees (Halictidae, Anthophoridae, Apidae). Open presentation of a common food reward (pollen) results in interactions with numerous pollinators¹⁰, a finding typical of plants with open flowers⁵.

Phylogenetic analysis of combined nuclear ribosomal and chloroplast DNA data sets, and mapping of pollination and morphological traits onto the molecular tree, indicate that Malagasy species of *Dalechampia* are descended from an ancestor pollinated by resin-collecting bees (Fig. 1). These results also indicate that *Dalechampia* colonized Madagascar from Africa (Fig. 1). This finding is further supported by morphological data and biogeographical considerations¹⁰.

But why did the Malagasy colonists 'abandon' efficient, specialized pollination by resin-collecting bees and switch to a generalized pollination system? It seems that resin-collecting megachilid bees, which are the only pollinators of *Dalechampia* in Africa⁹, failed to colonize Madagascar¹⁰. The ancestral *Dalechampia* colonists of

Madagascar were probably pollinated incidentally by other pollen-feeding insects. They subsequently adapted to the absence of their specific pollinators by losing the gland that secretes the resin reward and by effectively using diverse pollen-feeding insects as pollinators. These changes were sufficiently successful to allow secondary diversification on the isolated island of Madagascar.

W. Scott Armbruster

Institute of Arctic Biology, University of Alaska,
Fairbanks, Alaska 99775, USA

and Department of Botany,
Norwegian University of Science and Technology,
N-7034 Trondheim, Norway

e-mail: ffwsa@uaf.edu

Bruce G. Baldwin

Jepson Herbarium and Department of Integrative
Biology, University of California,
Berkeley, California 94720, USA

1. Futuyma, D. J. & Morena, G. *Annu. Rev. Ecol. Syst.* **19**, 207–233 (1988).
2. Thompson, J. N. *The Coevolutionary Process* (Univ. Chicago Press, 1994).
3. Smith, A. B. & Jeffery, C. H. *Nature* **392**, 69–71 (1998).
4. Givnish, T. J., Sytsma, K. J., Smith, J. F. & Hahn, W. J. in *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago* (eds Wagner, W. L. & Funk, V. A.) 288–337 (Smithsonian Institution Press, Washington D C, 1995).
5. Lanyon, S. M. *Science* **255**, 77–79 (1992).
6. Proctor, M., Yeo, P. & Lack, A. *The Natural History of Pollination* (HarperCollins, London, 1996).
7. Waser, N. M., Chittka, L., Price, M. V., Williams, N. M. & Ollerton, J. *Ecology* **77**, 1043–1060 (1996).
8. Armbruster, W. S. *Am. J. Bot.* **71**, 1149–1160 (1984).
9. Armbruster, W. S. & Steiner, K. E. *Am. J. Bot.* **79**, 306–313 (1992).
10. Armbruster, W. S. *et al.* *Natl. Geogr. Res. Expl.* **9**, 430–444 (1993).

The 'great secret' of chemistry's past

The Aspiring Adept: Robert Boyle and his Chemical Quest

by Lawrence Principe

Princeton University Press: 1998. 339pp. \$45, £32.50

D. M. Knight

We thought we knew where we were with Robert Boyle (1627–91), father of chemistry and uncle of the Earl of Cork. Boyle, so we believed, had no time for alchemy or other mysterious activities. He promulgated both a modern definition of a chemical element and a mechanical world-view. The world was a bigger version of the famous Strasbourg clock; matter was composed of corpuscles, all made of the same stuff, differently arranged. Boyle's book *The Sceptical Chymist* was reverently regarded as the beginning of rational chemistry, even if few actually read its long-winded dialogues.

Alas, just when we get our world picture clear, researchers come along and disturb it. Lawrence Principe has done just this in his provocatively titled *The Aspiring Adept*—like Geoffrey Keynes calling Newton *Last of the Magi* half a century ago. How could our Boyle have been, or have aspired to be, an 'adept'—the term for alchemists who have attained the 'great secret'?

A great deal of work has been done on the Boyle papers in recent years, particularly by Michael Hunter. The close study of the documents, and of associated published texts, has confronted us with a quite different Boyle. His lifetime was a period of tremendous religious and political upheaval. While he kept out of political storms, he was an important figure in the religious life of his time, supporting missions and biblical translations. As a good Protestant, he could not accept that saints might come between him and God. In the same way, he could not believe that Nature, in the form of a semi-divine demiurge, could come between his free and inscrutable God and the world. Boyle's preoccupations were typical of his time, and he impressed contemporaries as much by his piety as by his acuity.

Principe has approached Boyle from further back in history, from Basil Valentine, the fifteenth-century German monk whose name appears on chemical writings over a span of two centuries, rather than from later science—as a successor to earlier chymists rather than as a precursor to modern chemists. Careful reading of *The Sceptical Chymist* and other familiar writings of Boyle's indicate that he was not simply a mechanical philosopher who insisted, in a modern manner, that science was public knowledge and that experimental data should be published uncooked. Principe sees alchemy as part of the 'normal science' of the



seventeenth century, and *The Sceptical Chymist* as an attack only on vulgar alchemists: learned adepts escape criticism.

Principe prints as an appendix fragments from Boyle's *Treatise on Transmutation* (also in dialogue form, partly in Latin and partly in English) and also a very curious fragment of dialogue about converse with angels—a fascinating topic to our ancestors, where one needed to be sure that the angels were not fallen ones. In the fragments, Boyle reports on transmutations he and other reputable persons had witnessed. Certainly, they seem as well attested as other events in history that we do not doubt, though our chemistry does not allow us to believe them. Like the psychical researchers of the last century, Boyle was not the sceptic he sometimes thought himself.

We should note with Principe, however, that alchemy sometimes included a particular theory of matter, and that Boyle's corpuscles were readily compatible with transmutation. He was particularly struck with an experiment in which he seemed to have degraded gold to a base substance: his alchemy could go both ways, and he was impressed with the great effects that small quantities of potent reagents could produce.

Boyle had inherited wealth, some of which seems to have gone to support a mysterious con man, who promised him contacts with sages from the East, and alkahest, the universal solvent sought by the alchemists. But Boyle believed that it should be possible to live modestly by alchemy. Seeing the old statute against

it as a barrier to serious scientific research, he managed to get it repealed. References to alchemy are widely dispersed throughout his work and, although in some respects an advocate of openness, he was content to fall in with the secrecy of alchemy, sending letters in code, for example. Just as he allowed for God's action in a clockwork universe, so he was content with parallel (and seemingly incompatible) explanations in his chymistry.

The new Boyle is more interesting, because more filled out, than the old one. If both he and Newton took alchemy seriously, then we have to see chymistry as an ancestor of chemistry. We do not need to fret about this, however, any more than about our own remote and ape-like ancestors. In the late eighteenth century the alchemist Peter Woulfe FRS invented a wash bottle for gases, and attached to his apparatus little pieces of paper with prayers (we might have done the same, faced with his glassware). Michael Faraday was sympathetic to alchemy in principle, believing that there couldn't be so many irreducible elements. Ernest Rutherford sometimes described his work as "new alchemy".

Boyle the aspiring adept is in some ways further from us, in that foreign country of the past where they do things differently. But in other ways he is less remote, and Principe is to be congratulated on bringing him into a new focus. □

D. M. Knight is in the Department of Philosophy, University of Durham, 50 Old Elvet, Durham DH1 3HN, UK.

Eclipsed no more

Osmanli Astronomi Literaturu Tarihi
(History of Astronomy Literature during the Ottoman Period)

edited by Ekmeleddin Ihsanoglu
International Research Centre for Islamic History, Culture and Arts: 1997. 1,146pp.
(two volumes). \$100

Ziauddin Sardar

Ottoman science has long been considered a poor descendant of Islamic science. It has been largely neglected, and even suppressed, partly because of its association with the 'decline' of Muslim civilization and partly because of European perceptions of the Ottoman state as an enemy. The project on 'scientific literature in the Ottoman period', sponsored by the Organisation of Islamic Conference and undertaken by the International Research Centre for Islamic History, Culture and Arts (IRCICA), is designed to highlight the achievements of Ottoman science and restore it to its rightful place.

It is a massive enterprise, inspired to some extent by Joseph Needham's work on Chinese science, and running into several volumes. Its nature and scope, as well as the problems involved in such an undertaking, are well illustrated by the first two volumes, on Ottoman astronomy.



Star wars: the Ottomans would use astronomical events, such as an eclipse, to plan military campaigns.

The problems begin with definition. What constitutes Ottoman science? The Ottoman dynasty held power for 600 years, an extraordinary achievement in a region where states and empires emerged and then disappeared in a few generations. During that period the extent of Ottoman wealth and power changed enormously. At the beginning of the thirteenth century, the Ottomans held no more than a small area of northwest Anatolia. But by the end of the fifteenth century they were an important regional power, ruling most of modern Turkey and a large part of the Balkan peninsula. During the sixteenth century they became masters of a vast, multinational empire that stretched from Slovakia to Nubia and from Algiers to the Caucasus.

So where does Ottoman science begin and end? The IRCICA researchers solved this problem by focusing on scientists who lived, or spent part of their lives, in Ottoman lands between 1299 and 1923.

The location of the literature presents another problem. Ottoman scientific treatises are to be found not just in all the many countries that were once part of the Ottoman Empire but also in Western Europe and the United States. In this regard, the researchers left no library, no archive, no museum, unexplored.

The result is a monumental achievement. It not only provides us with a true picture of the extent of Ottoman scientific activity, but also turns the standard view on its head. The conventional view is that the empire reached its peak around 1600 and then spent the next 300 years generally declining, when science disappeared and technology evaporated. *Osmanli Astronomi* shows that, far from disappearing, science was very much alive in the Ottoman Empire right up to the eighteenth century, when it

shifted towards learning and assimilating European sciences through translations and adaptations.

The survey of the literature provides information on more than 600 original scientific minds, including those whose exact period is unknown, with details of where their work is located. Each entry gives a brief biography of the scientist, an outline of his scholarly career, a list of his works on astronomy and of any secondary literature relating to the author, and an indication of the language in which he wrote (normally Arabic, Turkish or Persian). A brief history and full bibliographical information for each work is also given.

Much of this literature consists of theoretical and practical astronomy. Included are works exploring orbits of the planets, the relationship between astronomy and Islam, zîjes-tables (mathematical tables for calculating times of sunrise and sunset, planetary positions, eclipses and so on), calendars and astronomical instruments. The introduction provides a good insight into Ottoman astronomical institutions such as the Istanbul Observatory and the timekeeping institution Kandilli Rasathanesi. An extensive bibliography provides a guide to recent literature on these subjects.

Osmanli Astronomi is a truly prodigious work that will be celebrated by historians of Islamic science everywhere. It is not without its faults — there are poor translations and numerous typographical errors — but these are largely insignificant given the vast scope of the undertaking. The next volumes in the project will focus on mathematics and geography and will undoubtedly continue to rewrite the history of Ottoman science. □

Ziauddin Sardar is editor of *Futures*. He is at 1 Orchard Cafe, London NW9 6HE, UK.



Sun signs: Pisces seen through Ottoman eyes.

Dogged by controversy

The Brown Dog Affair

by Peter Mason

Two Sevens Publishing, 30 Wynter Street,
London SW11 2TZ: 1998. 118pp. £5.50

Lethal Laws: Animal Testing, Human Health, and Environmental Policy

by Alix Fano

Zed: 1998. 256pp. £37.50, \$55 (hbk), £12.95,
\$19.95 (pbk)

John Galloway

On the face of it, these two books are about the way animals are used by scientists. They tell us as much, however, about people's frustration with authority — government, professional and commercial — and their sense of being denied the chance to influence events. *The Brown Dog Affair* is a revealing piece of social history which shows this rather clearly. *Lethal Laws* is a relentless attack on the use of animals for toxicity testing, and a call to governments to ban testing on live animals

The Brown Dog Affair began in 1903 when Dr (later Sir) William Bayliss of University College London's physiology department issued a writ for slander and libel against Stephen Coleridge, honorary secretary of the National Anti-Vivisection Society.

At a large public meeting at St James's Church in London's Piccadilly, described by the *British Medical Journal* as "carefully stage-managed", Coleridge had savagely attacked Bayliss over a demonstration he had given to medical students. Bayliss had cut open a dog's neck to expose the salivary gland, which was stimulated electrically, in an unsuccessful attempt to show that the pressure at which saliva is secreted is greater than blood pressure. A "terrier-type" dog was used — the brown dog of the title.

Coleridge accused Bayliss of breaking the law by twice contravening the 1876 Cruelty to Animals Act, for which the National Anti-Vivisection Society had campaigned. After the demonstration the dog had to be put down. Coleridge did not leave the matter there, and in his rabble-rousing speech he invited the public to condemn vivisection on principle.

By naming Bayliss, Coleridge guaranteed that the case came to court. Bayliss won and was awarded £2,000 damages, a huge sum for the day. But, as Coleridge intended, public interest and sympathy for the dog was aroused. The damages and costs were paid by public subscription, the story was taken up in the national press, and questions were asked in the House of Commons.



Canine cause: monuments to the brown dog have done little to stop animal experimentation.

After the trial, a plan was hatched to keep the story alive. Louisa Woodward, a wealthy Londoner given to charitable causes, paid for an elaborate granite drinking fountain surmounted by a bronze statue of 'the dog'. In 1906 she persuaded Battersea Council, then strongly anti-establishment, to erect it in a local recreation ground. The fountain's inscription, although factually correct, was obviously inflammatory and — University College claimed — libellous. It read, "In memory of the brown terrier dog done to death in the laboratories of University College", and continued in a similar vein.

As a way of keeping the controversy alive, the memorial succeeded famously. Medical students demonstrated in protest and attacked the monument. The *British Medical Journal* thundered against it. United, perhaps, by a sense of being underdogs themselves, trades unionists and some representatives of the suffragette and early labour movements rallied in support of the monument. It was seen as a symbol of political progress towards greater justice in the social order and a protest against the establishment, as epitomized by the medical practitioners of the day. If proof were needed, when the Conservatives were returned to power in Battersea in 1910, the fountain was removed and later destroyed. *The Lancet* breathed a sigh of relief, announcing that the monument had at last been overcome by "considerations of common sense".

Though dead, the monument refused

to lie down. In 1985 a new statue with a new inscription, which incorporated the original, was unveiled in Battersea Park, paid for by the National Anti-Vivisection Society and the British Union for the Abolition of Vivisection. The *British Medical Journal* and University College predictably condemned it. This, too, disappeared in 1992, but in 1994 emerged again, although somewhat less prominently sited.

It is striking that the monument, or the idea behind it, survived intermittently for nearly 90 years and still had the power to provoke the medical establishment. However, the number of animals used in experiments in the United Kingdom rose from 19,000 each year to more than 3 million in the 90 years following the brown dog's death, a tide the statue did nothing to stem.

It is evident from *Lethal Laws* that most research using live animals is not now done by physiologists trying to explain how the body works. Problems of structure and function have been driven down to cellular and molecular levels. Rather, the use of animals is the result of legislation about public health, and health and safety at work — to protect people ever more dependent on new and 'improved' chemical products, drugs, pesticides, cosmetics and detergents. The results of these tests are generally uncertain and their implications probably more so. In practical terms testing is no more than a crude screening process; whether the metabolic similarities between animals and humans are more significant than their differences is often a matter of luck. Fano is unhappy with this inbuilt uncertainty, drawing the conclusion that if animal testing is not foolproof it can play no part in protecting the public. She recounts the story of thalidomide, where animal testing failed to identify a novel side-effect that had disastrous consequences. This is a powerful and emotive example, but not one from which it is easy to generalize. And it is naive to believe that chemicals can be straightforwardly divided into beneficial and harmful — there is always a trade-off. All drugs probably cause unwanted side-effects for someone, somewhere.

Like Coleridge at the beginning of the century, Fano advocates a recourse to law, together with a campaign to move the issue into a wider social domain. For her this involves finding allies in the Green movement, while acknowledging that the 'Greens' think the animal lobby sentimental. The rationale she seeks to persuade them with, therefore, is that animal testing in isolation is simplistic and encourages unwarranted optimism about the effect of any chemicals on the

environment. Her legal angle is interesting, namely that extrapolating the findings of animal testing to people is unscientific and fraudulent. This might be a legal long shot but, like the Brown Dog Affair, it could make a good publicity stunt, given the media interest in scientific fraud.

The Brown Dog Affair is an eminently readable book, giving a strong flavour of the times. Peter Mason gives a balanced but humane view, although it would have helped to have a better picture of the medicine and surgery of the day. There is a hint at one point that poorer people were themselves used as 'guinea pigs' for new treatments — it would have been interesting to know more.

Lethal Laws does not aim to achieve a balance — arguments in favour of animal testing are briefly rehearsed, but only to be knocked down. The real question is how hard the evidence is to support Fano's claims. She acknowledges that many of the examples she gives date back 20 years. Also, she claims that pollution is the cause of a cancer epidemic, when the most significant reasons for its increasing incidence in developed countries are that people are living longer and women smoking more. Both books are about confrontation, born of frustration. But more tends to be achieved by pragmatic compromise and horse trading. □

John Galloway is in the Team Development Unit, Eastman Dental Hospital, 256 Grays Inn Road, London WC1X 8LD, UK.

Grub's up!

The Eat-A-Bug Cookbook

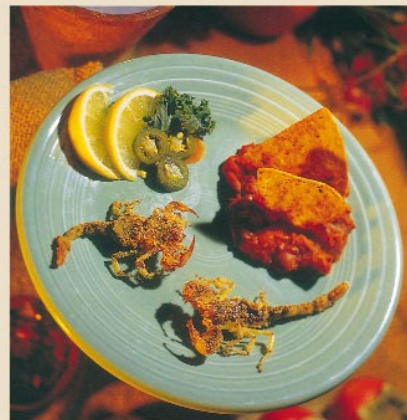
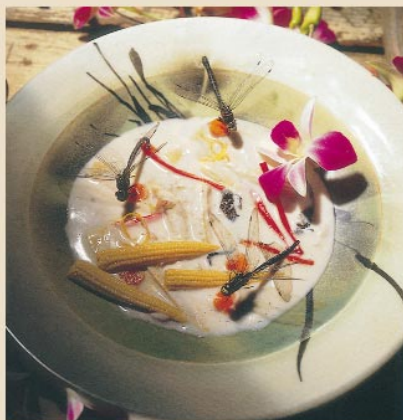
by David George Gordon

Ten Speed Press: 1998. 140pp. \$12.95 (pbk)

Helen Phillips

Entomophagy — eating insects — is nothing new. Most of the world's indigenous peoples include insects on the menu. But somewhere along the line, the population of Europe and the United States has lost the taste for a large proportion of the phylum Arthropoda. Why should we relish crustaceans, yet reject the culinary delights of their protein-rich, cholesterol-free, vitamin-laden six-legged or eight-legged cousins?

Other entomophages have attempted to right this anomaly. In 1885, Vincent M. Holt published a persuasive text entitled *Why Not Eat Insects?*, in which he attempted to convince England's starving peasantry that it would be much better fed if it were willing to embrace the locally collected invertebrate fauna. More recently, Ronald L. Taylor's inspired work *Butterflies in My Stomach*, and the accompanying volume *Entertaining with Insects* tried to appeal to



Insect cuisine: From top left clockwise... Sheesh! Kabobs, Cream of Katydid soup, Scorpion Scaloppine and Sky Prawns.

the open-minded 1970s generation. As these authors evidently failed to change our habits, it would seem fair for another to try to convince us of the delights of an insect-rich diet.

In *The Eat-A-Bug Cookbook*, David George Gordon makes a fine attempt (he uses 'bugs' in its widest sense, to include anything that creeps, crawls, hops or wriggles). There is a very interesting section on government-approved entomophagy, which might make your skin crawl more than the raw ingredients do.

The US Food and Drug Administration has officially sanctioned permissible degrees of insect damage and infestation, including specified numbers of eggs, immature and adult insects, or their parts, that are allowed in various foods. It is acceptable if there are two or three fruitfly maggots in 200 grams of tomato juice, and as many as 100 insect fragments are permitted in 25 grams of curry powder. And if you've ever wondered why fig seeds stick between the teeth, then you'll appreciate knowing that up to 13 insect heads are allowed in 100 grams of fig paste. Not to mention cochineal, a red food additive, made from the pulverized remains of the bug (this time the word is used in its true sense) *Dactylopius coccus* — 70,000 bugs in every pound.

This useful information is not intended

to put us off our salad, or to scare us into making hand-lens inspection a routine part of food preparation, but instead to convince us that eating insects is officially OK. To whet your appetite, here are some of the delicacies on offer. 'Three bee salad' and 'Nine ways to turn the tables on household and garden pests' were two of my favourites, closely followed by 'Party pupae', 'Cream of Katydid soup' and 'Pest-o'.

For the inexperienced entomophagist, there are heaps of useful cooking tips, nutritional information, historical accounts, preparation techniques (how to disarm a scorpion or a bee, and how to prepare a cricket before it hops off), serving suggestions, and hints about biological suppliers or harvesting from the wild. There is even a description of how to create the cricket equivalent of Club Med, should you wish to keep fresh ingredients on hand. The anatomical details and 'choice cuts' section, however, was just a little more information than I really needed.

No self-respecting recipe book would be complete without its recommendations of the best wines to accompany the meal — apparently you can't beat champagne. The best advice was to take two bottles, one to eat with the meal and one to fortify the chef. □

Helen Phillips is a science writer at Nature.

Chaotic topography, mantle flow and mantle migration in the Australian–Antarctic discordance

David M. Christie*, Brian P. West*†, Douglas G. Pyle‡ & Barry B. Hanan‡

* College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331-5502, USA

† Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

‡ Department of Geological Sciences, San Diego State University, San Diego, California 92182-1020, USA

Oceanic crust formed over the past 30 million years at the Australian–Antarctic discordance (AAD) is characterized by chaotic sea-floor topography, reflecting a weak magma supply from an unusually cold underlying mantle. During the past 3–4 million years, however, a source of increased magma supply, coinciding with the known Indian–Pacific mantle isotopic boundary, has propagated into the eastern AAD, displacing the chaotic terrain and replacing it with normal sea floor. Pacific mantle reached the eastern boundary of the AAD at least 7 million years ago, but it was not until 3–4 million years ago that lavas derived from Pacific mantle were first erupted within the AAD. This long hiatus, combined with the ridge–transform geometry across the AAD boundary, constrains the locus of mantle migration to a narrow, relatively shallow region, directly beneath the spreading axis of the Southeast Indian ridge.

The deepest parts of the Southern Ocean and, indeed, of the global mid-ocean-ridge system, lie within the AAD, centred on the Southeast Indian Ridge (SEIR) between 115° E and 130° E (ref. 1; Fig. 1). The unusually great depths within the AAD result from an unusually cold underlying mantle that yields a low melt supply to the spreading plate boundary, and forms relatively thin oceanic crust^{2–4}. Across the eastern boundary of the AAD, marked changes in virtually all aspects of crustal accretion have been the foci of a number of recent geophysical and geochemical studies^{4–8}. Also present within the eastern AAD is a distinctive boundary between “Indian-type” and “Pacific-type” upper-mantle provinces^{9,10}. Here we present two important observations from a recent bathymetric, geophysical and petrological survey of the eastern boundary region of the AAD, and consider their implications for long-term mantle dynamics of the region.

(1) New multibeam bathymetric data (Fig. 2) show that the deepest parts of the AAD have, for at least 30 Myr, been characterized by a distinctive, chaotic sea-floor terrain that reflects a predominance of tectonic over magmatic extensional processes.

(2) Geochemical analyses of new dredge samples from off-axis sea floor (0–7 Myr age) confirms that the Indian–Pacific isotopic boundary has migrated westwards into the AAD during the past 3–4 million years. This migration coincides with a discrete increase in magma supply that has propagated across the easternmost AAD, displacing the chaotic terrain and creating a ‘V’-shaped, normal abyssal-hill terrain in its wake.

These observations refine our understanding of the history of mantle migration and its relationships to the unique dynamic, geophysical, and petrologic features of the AAD during the past few million years. For the longer term, although we cannot yet uniquely unravel the history of the Indian–Pacific mantle boundary, the recent (3–4 Myr) migration within the AAD enables us to focus on two competing hypotheses. The recent migration represents either the culmination of a long-term (30–50 Myr) westward mantle migration from the Pacific region into the Southern Ocean, or a relatively minor perturbation in the position of a boundary that has been fundamentally linked to the AAD throughout its history. Here we consider these two hypotheses in the light of newly available geological evidence.

Tectonic boundaries of the eastern AAD

Our bathymetric maps (Fig. 2a, b) reveal two distinct styles of sea-floor morphology in the eastern AAD region. East of the 127–128° E (bounding) transform, ‘normal’ sea floor is dominated by typical axis-parallel (east–west) abyssal hills. Within the deepest part of the AAD, west of the 126–127° E transforms, the sea-floor terrain

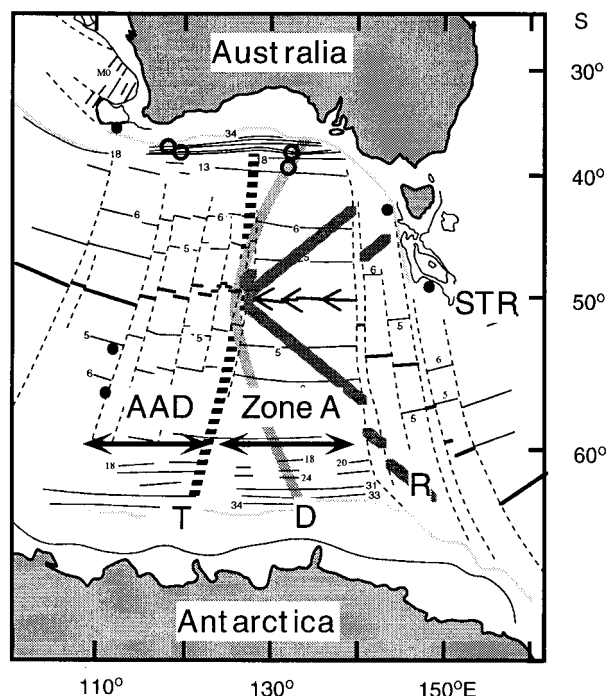


Figure 1 Regional setting of the Australian–Antarctic Discordance (AAD). The Southeast Indian Ridge (SEIR) is shown as a heavy black line, the main fracture zones as dashed lines, and selected magnetic anomalies as thin black lines. Broad shaded curves illustrate alternative configurations of the Indian–Pacific oceanic mantle isotopic boundary; T, eastern transform boundary of the AAD; D, residual depth anomaly; R, boundary created by rapid mantle migration; STR, South Tasman rise; closed circles, Deep Sea Drilling sites³⁰; open circles, dredge sites of Lanyon *et al.*³¹.

is unusually chaotic, dominated by irregular shallow blocks that are separated by irregular, deep valleys and typically bounded by orthogonal, axis-parallel and axis-perpendicular scarps up to 1,600 m high.

Between these transforms, in the easternmost AAD segment (B5), 'normal' sea-floor terrain occupies a west-pointing, 'V'-shaped region that is symmetrical about the spreading axis (Fig. 2). This terrain has progressively displaced the older chaotic terrain, beginning in the east 3–4 Myr ago and migrating westward, reaching the western boundary of segment B5 only recently. The two terrains are separated by a transitional zone, shown as a ruled area in Fig. 2c. The inner boundary of this zone separates the uniform abyssal-hill topography within the 'V', from rougher, and in places shallower, sea floor to the west (Fig. 2a, b). The outer (older) boundary is more difficult to delineate. It separates the transitional zone, within which the terrain is irregular but still dominated by axis-parallel fabric, from the completely 'chaotic' topography farther from the axis.

The overall 'V'-shape of the transition region and its boundaries implies an average migration velocity close to 40 mm yr^{-1} relative to the plate boundary, consistent with measured migration rates of propagating rifts farther east and with recent numerical modelling (see below). The 'V'-shape is also reminiscent of pseudofaults produced by propagating rifts elsewhere¹¹, but in this case there is little evidence for the usual hallmarks of rift propagation, relocation and reorientation of the spreading axis. Rather, the progressive establishment of normal sea-floor accretionary processes across segment B5 implies the introduction of a persistent, stable magma supply into the AAD and its east-to-west propagation across segment B5 during the past 3–4 Myr.

Amagmatic extension within the AAD

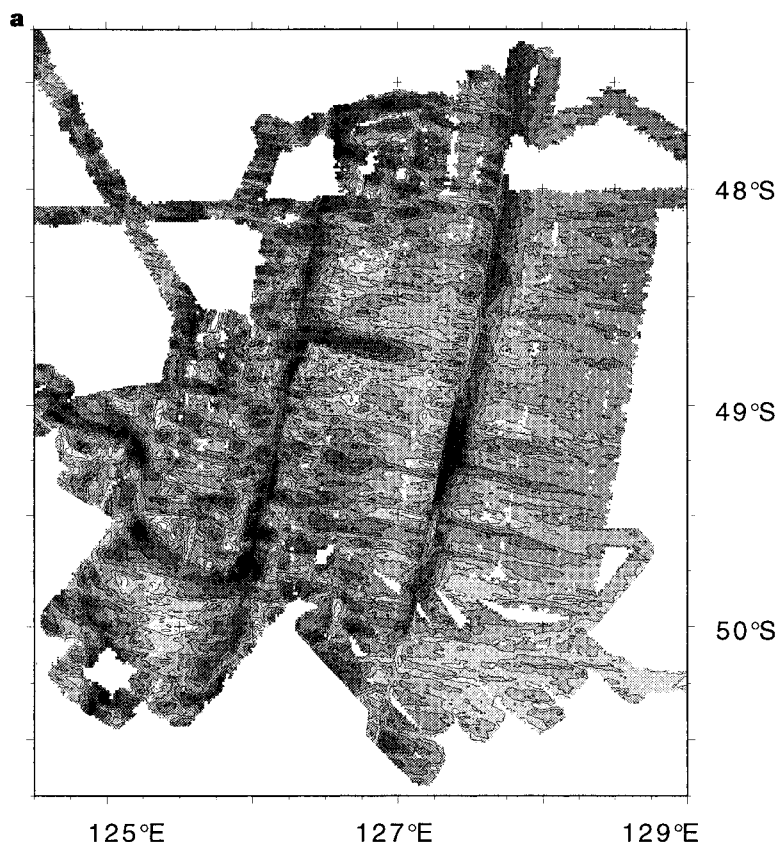
Important tectonic features of the chaotic terrain include the

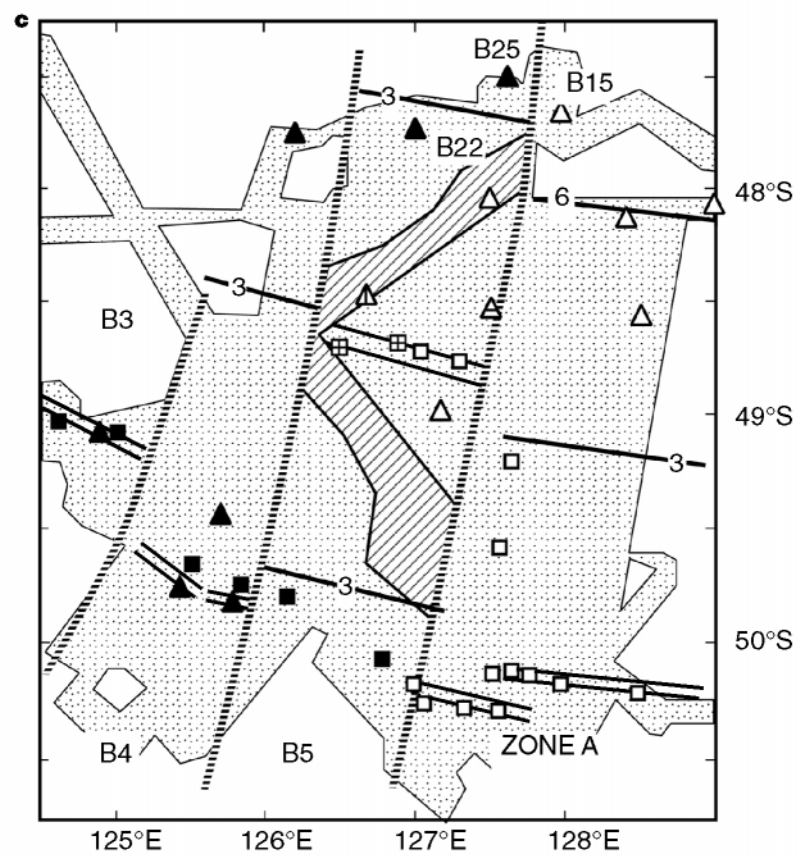
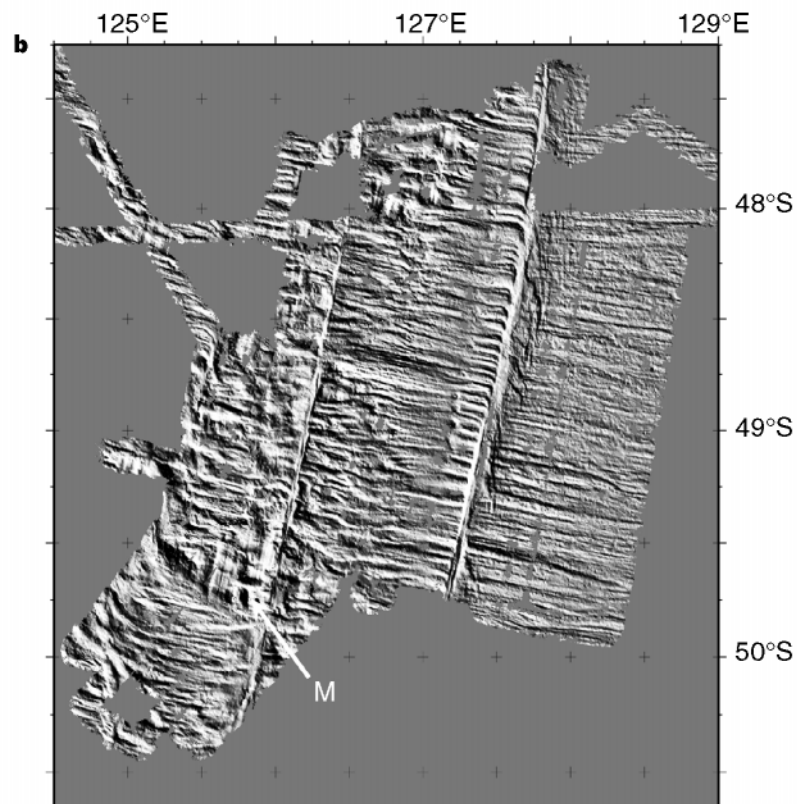
absence of persistent, axis-parallel tectonic fabric, the common presence of isolated, elevated blocks and of axis-normal lineations, and the close, but apparently random, association of both axis-parallel and axis-normal lineations. These features are most clearly revealed when an oblique, false illumination is applied to the sea-floor topographic data (Fig. 2b). Also revealed by the oblique illumination are sets of parallel, closely spaced, axis-normal lineations. These are "megamullions", recently recognized along the Mid-Atlantic Ridge as tectonic striae on the footwalls of long-lived, low-angle detachment faults^{12,13}.

Dredging of chaotic sea floor has recovered a varied igneous rock suite. Pillow basalts occur throughout the region, while diabase, gabbro and serpentized peridotite have been recovered from several elevated blocks and from axial valley walls in segment B3¹⁴. Calculated average sea-floor densities are unusually high ($\sim 3,050 \text{ kg m}^{-3}$) across the chaotic terrain of segments B4 and B5 (Fig. 3), strongly suggesting that the densest rock types, serpentized ultramafics, predominate at or near the sea floor. By contrast, calculated sea-floor densities are normal ($\sim 2,600 \text{ kg m}^{-3}$) in zone A. Similar high densities occur over a "megamullion" near 30°N on the Mid-Atlantic Ridge^{12,15}.

Although many of its features occur individually elsewhere, especially along amagmatic segments of the Mid-Atlantic Ridge^{16–20}, the chaotic terrain of the AAD appears to be unique in its areal extent and longevity, its apparent independence from axial segmentation, and its occurrence at a relatively high spreading rate ($\sim 37 \text{ mm yr}^{-1}$, half rate). The persistence of this terrain, and its association with the axis of the depth anomaly, indicate a 'permanent' magma deficiency associated with the persistent cold mantle beneath the AAD. Comparable sea-floor terrains may exist in the equatorial Atlantic, but only close to fracture zones and confined within individual segments¹⁹. High-strain, disorganized topography with abundant ultra-mafic exposures also occurs along the extremely slow spreading Southwest Indian^{21,22} and American-

Figure 2 Bathymetry of the eastern boundary region of the AAD to a maximum sea-floor age of $\sim 7 \text{ Myr}$. **a**, Gridded and contoured multibeam bathymetry. Contour interval is 250 m. The primary data source is our new Seabeam 2000 survey, supplemented by a 1998 SeaMARC II survey⁶. The extent of SeaMARC II coverage can be estimated by comparison with **b**, from which it is excluded. **b**, Seabeam 2000 bathymetry illuminated from direction 065° . This serves to highlight structural trends on the sea floor, independent of depth. SeaMARC II data included in **a** are excluded here due to noise, reducing the areal coverage. **M** shows the location of a large 'megamullion' discussed in the text. **c**, Overlay diagram showing locations of dredged samples and essential tectonic features relative to the mapped area of **a**. Older, mainly axial¹⁰, dredges are shown as squares, new off-axis dredges as triangles. Symbols are filled according to their mantle sources as inferred from Fig. 5: Open symbols, lavas derived from a Pacific-type mantle source; closed symbols, those derived from an Indian-type mantle source; lines within symbols, a transitional Pacific source³⁰. Double lines mark the axis of spreading. Numbered single lines are isochrons. B3, B4, B5 are individual segments within the AAD. Zone A refers to the region east of the AAD¹.





Antarctic ridges²³, but details of its morphology and its geographical extent are not yet known.

Although the high calculated sea-floor densities indicate a predominance of ultramafic rock types at the sea floor, the widespread recovery of pillow lavas from the chaotic terrain, especially from the axis and other deep valleys, indicates significant, possibly episodic, eruptive activity. Collectively, these lavas differ from those of 'normal' sea floor in their unusually high geochemical diversity^{10,14}. Indeed, samples from two axial dredge sites in chaotic topography of segments B3 and B4 encompass

almost the entire known range of $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios for the entire Indian Ocean spreading system^{10,14}. These observations suggest that ambient mantle heterogeneities are incorporated and preserved in diverse lavas that form an irregular veneer over a dominantly plutonic basement²⁴.

Chaotic sea-floor terrain can also be recognized in older (> 10 Myr) sea floor to the north of the SEIR, wherever the sparse multibeam swaths cross the residual depth anomaly (Fig. 4). Both the chaotic sea floor and the axis of the depth anomaly cross the eastern bounding fracture zone of the AAD near $44^\circ 30' \text{S}$. This is a crucial observation. Because the essentially shallow tectonic processes that produce the chaotic topography have been able to cross the fracture zone, independent of the axial segmentation, they must be initiated by deeper-seated changes in one or more properties of the mantle. Because the depth anomaly is presumed to be the locus of minimum mantle temperature along the axis of the SEIR, its close association with the chaotic terrain further suggests that chaotic terrain has formed in response to a long-term magma supply deficiency associated with cold mantle beneath the AAD.

The AAD as a geochemical boundary

Within the eastern AAD, the Indian–Pacific mantle isotopic boundary occurs over a few tens of kilometres at, or close to, the transform between segments B4 and B5¹⁰ (Fig. 2c). Although transitions between comparable, ocean-scale mantle isotopic domains have been recognized elsewhere along the mid-ocean-ridge system^{25–29}, they are typically diffuse, occurring over more than 100 km. Westward migration of the Indian–Pacific isotopic boundary was first inferred by Pyle and others¹⁰ from the Indian Ocean isotopic signatures of two off-axis samples in segment B5^{10,30,31}.

Lead isotope ratios for basalt glasses from 16 new off-axis dredge sites are plotted in Fig. 5 in relation to the previously defined¹⁰ 'Indian' and 'Pacific' fields for axial lavas. Using these fields, the mantle source type for each off-axis sample is determined and its map location is identified in Fig. 2c. The 'Indian' character of dredges B22 and B23 to the north of the B5 axis (Fig. 2c) corroborates the presence of Indian mantle beneath segment B5 at 3–4 Myr, while the westernmost Pacific-type samples lie within the relatively shallow transition zone between the normal and chaotic terrains of segment B5. The off-axis trace of the Indian–Pacific boundary within B5 must, therefore, lie close to the outer margin of the transition zone, confirming the close association between the transition from chaotic to lineated topography and migration of the underlying mantle.

East of the AAD, all analysed samples from seafloor of 0–7 Myr age are clearly derived from Pacific mantle (Fig. 2c). The northernmost of these samples is a glass fragment preserved in the core of a manganese nodule from ~7-Myr-old sea floor (dredge B15). Because this sample is not actually *in situ*, it must be interpreted with caution. But even a detached, rounded nodule seems unlikely to have been transported even a few kilometres across rough terrain. It seems reasonable, therefore, to conclude that dredge B15 defines ~7 Myr as the youngest possible age for eruption of Indian-type samples east of the AAD (Fig. 2c). This key observation constrains the possible timing of the hypothesized, long-term, westward migration of Pacific mantle across zone A (see discussion below). Indian mantle that was present east of the AAD 30 Myr ago³¹ must have been displaced by migrating Pacific mantle before ~7 Myr ago, at least 3–4 Myr before the first known eruption of Pacific-type lavas within the AAD. These observations of the location and surface expression of the mantle boundary provide important, independent constraints on the geometry of asthenospheric flow into and beneath the AAD.

Cool mantle and the origin of the AAD

The AAD and its associated depth anomaly are long-lived manifes-

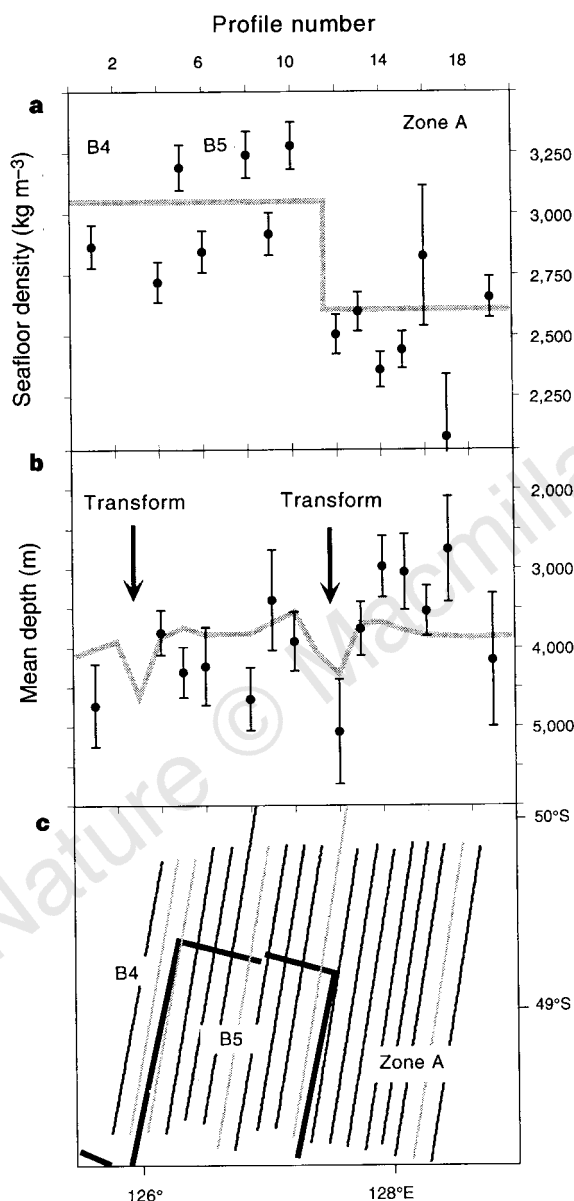


Figure 3 Average sea-floor densities and depths calculated from gravity data along profiles on either side of the eastern AAD boundary. **a**, Sea-floor densities for individual profiles. Overall averages for the AAD and zone A are shown by the grey line. **b**, Best-fit depths for individual profiles. Error bars denote 95% confidence intervals assuming a strictly gaussian distribution. The grey line connects average observed depths for each profile. **c**, Profile locations (parallel lines) relative to the plate boundary through Zone A, B5 and B4. Only profiles shown in black were used in the analysis. Profiles (shown in grey) for which the data are not linear for wavelengths between 10 and 100 km and those that cross transform fault traces were not used. Details of how we obtained sea-floor density and depth from gravity data are given in Methods.

tations of cool upper mantle that has persisted beneath the region at least since separation of Australia from Antarctica. The fracture zone extensions of the main AAD transform faults stretch from the Australian to the Antarctic continental margin (Fig. 1). The residual depth anomaly (the off-axis trace of the deepest part of the spreading centre) also spans the ocean basin. It is roughly symmetrical about the spreading axis and cuts obliquely across the eastern AAD fracture zones, indicating slow ($\sim 15 \text{ mm yr}^{-1}$) westward migration of its locus in the sub-lithospheric mantle, relative to the axis of spreading³. This slow migration of the depth anomaly is in marked contrast to the more rapid mantle flow ($\sim 40 \text{ mm yr}^{-1}$) inferred from the recent migration of the isotopic boundary across B5 and from active ridge propagation along the SEIR towards the AAD.

This apparent contradiction can be resolved by careful consideration of mantle flow patterns and their sea-floor manifestations in both mantle (absolute) and plate (relative) reference frames. Regardless of their origin, cool upper-mantle temperatures beneath

the AAD create strong, regional thermal and viscosity gradients. These gradients lead, in turn, to strong, converging, shallow asthenospheric flows beneath the SEIR³². Because the depth anomaly coincides with the null region at the convergence of these two subaxial flow streams, it will migrate at a relatively low velocity in response to any imbalance between the converging flows. There are several theories concerning the nature and origin of the mantle cold region and, hence, of the depth anomaly.

Mantle cold spot. A cold point-source which is in some sense the inverse of a hotspot has been invoked by several authors^{33–35}. However, a point source fixed in a mantle reference frame could not have created the depth anomaly. Such a ‘coldspot’ must migrate northwards, in order to maintain its position beneath the spreading axis³. West⁸ suggested that a migrating coldspot could result from trapping of a fragment of sub-continental mantle beneath the SEIR.

Convective downwelling. It has frequently been postulated that the AAD overlies a north–south orientated, convective downwelling

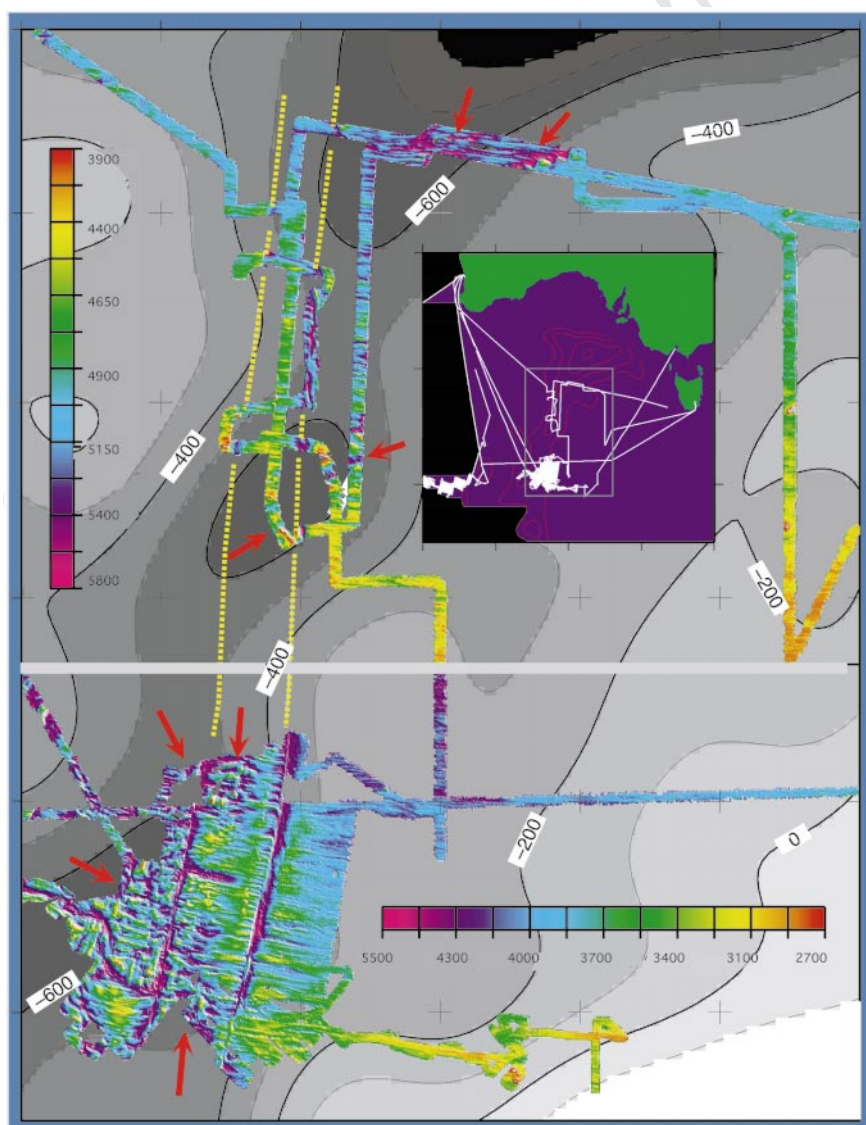


Figure 4 Results of multibeam bathymetry, illuminated from direction 065°, for part of the AAD and the region to the northeast, extending to sea floor as old as 30 Myr. We note that the colour scale is not uniform across the area. Separate colour scales, indicating depth in metres, are used north and south of 26.5°S in order to accommodate the normal off-axis depth increase, and to still highlight the extreme local depth variability associated with the chaotic topography. Contours

show the residual depth anomaly (contour interval, 100 m; recalculated by B.P.W., using the new bathymetry data). Red arrows indicate off-axis regions of chaotic topography and emphasize its association with the depth anomaly. Yellow dashed lines are approximate traces of important fracture zones. Inset, regional location map showing survey tracks and depth anomaly contours.

location of the isotopic boundary at the eastern edge of (rather than within) the chaotic terrain seems more consistent with the recent arrival of Pacific mantle from the east. If the boundary were fixed beneath the depth anomaly, one might expect to find both Pacific and Indian lavas within the chaotic terrain.

The sparse available geochemical data have also been taken to support a hypothesis of long-term migration. Lanyon and others³¹ recovered samples with Indian-type isotopic signatures from crust ~38 and ~45 Myr old, east of the AAD (Fig. 1), while Pyle and others³⁰ identified samples with Indian-type isotopic signatures from deep-sea drill holes close to Tasmania (Fig. 1). In fact, the Lanyon *et al.* samples lie within the depth anomaly, east of the fracture zone boundary of the AAD. They discount a third, possible, 'no migration' hypothesis, in which the isotopic boundary is tied in the long-term to the transform boundary of the AAD³⁰, but they do not distinguish between the short- and long-term hypotheses. Taken at face value, the samples of Pyle *et al.* are consistent only with long-term rapid migration; but recent discoveries of basalts with Indian-type isotopic signatures at spreading centres far from the Indian Ocean^{23,40,41} require that this conclusion be treated with caution until a clear spatial connection is established by further sampling. Discrimination between the two hypotheses will not be possible without a more systematic sampling of older sea floor. Appropriate sampling is planned for the Ocean Drilling Program in 1999.

Geometry of asthenospheric flow into the AAD

The shape of the isotopic boundary, as recorded at the sea floor by lavas erupted along a linear spreading axis, will not normally mimic the geometry of the boundary in the shallow mantle. The sea-floor boundary is shaped by the interaction of several parameters, including: the rate and direction of spreading; the rate of mantle migration; the geometry of the mantle boundary; the dynamics of mantle upwelling; and the dynamics of the melting region. When a migrating mantle front encounters a transform, its expression in sea-floor lavas will be further complicated by the ridge–transform–ridge geometry. In most cases, the first arrival of the mantle boundary beneath the upstream ridge–transform intersection will differ significantly from its first emergence beneath the downstream spreading segment (compare Fig. 6a with Fig. 6b, c). This time difference determines the offset pattern of the sea-floor boundary across the transform.

We now consider how our geochemical data can be used to constrain the possible kinematics of a Pacific mantle front under the rapid-migration hypothesis. To understand the relationships between such a mantle front and its surface expression, we begin with two simple hypothetical examples. Figure 6a shows the special case of a discrete, vertical, planar mantle front, striking parallel to the transform. In this special case the arrival of the mantle front at the transform and its emergence on the downstream side are simultaneous, and the offset distance from the sea-floor boundary to the spreading axis will be equal on either side of the transform. For any other geometry, there will be a significant time difference between upstream arrival and downstream emergence, and the offset patterns will be more complex. In the case of a mantle front with a curved leading edge (Fig. 6b), there will be a hiatus between the arrival of the sea-floor boundary at the transform and its first appearance on the downstream segment. The length of this hiatus increases with the curvature of the mantle front. Conversely, the length of an observed hiatus can be used to constrain the shape of a migrating mantle boundary.

For a simple, dipping mantle front³², similar arguments can be applied if one assumes that the location of the boundary at the surface is controlled by its location at some critical depth, perhaps that of the solidus. Such a depth will probably change gradually as mantle temperature changes, but drastic changes seem unlikely on the scale of this discussion.

In the case of the AAD, the absence of Indian-type samples from the sampled (0–7 Myr) area of zone A (Fig. 2c) implies that a westward-migrating Indian–Pacific isotopic boundary cannot have intersected the eastern transform of the AAD more recently than ~7 Myr ago. This implies a hiatus of at least 3–4 Myr between the arrival of Pacific-type lavas at the transform and their emergence at the eastern end of the B5 axis. Such a long hiatus can most simply be interpreted in terms of a mantle front that is strongly curved, as in Fig. 6c, rather than planar or gently curved (Fig. 6a, b). Other geometries may seem possible, but none seems plausible, given the geological and tectonic setting of the AAD and existing evidence for along-axis mantle flow. The strong curvature inferred for the mantle front is, in essence, equivalent to the confinement of mantle flow to a relatively narrow, low-viscosity region beneath the axis of spreading, consistent with a variety of regional numerical models that show a focusing of along-axis flow in the shallow asthenosphere beneath the spreading axis^{8,32}.

A second important constraint on Indian–Pacific boundary migration is the absence of Pacific-type lavas from segments B4 and B3 (Fig. 2c). In the simple model of Fig. 6c, mantle migration might be expected to continue undiverted beyond the transform, in which case Pacific mantle should now be present beneath B4 and probably B3. The absence of Pacific-type lavas from these segments (Fig. 2c) can be interpreted in terms of two complementary processes: (1) physical obstruction or diversion of westward mantle flow by the eastern AAD transform, and (2) northward migration, in an absolute reference frame (and relative to the Antarctic plate), of the spreading centre, effectively leaving behind any Pacific-type mantle that crosses the transform into B5⁴².

The absence of Pacific lavas from B4 precludes any possibility of migration along a broad mantle front that is insensitive to plate-boundary geometry, and reinforces the notion that mantle migration is strongly controlled by spreading geometry, especially by thermal gradients in the shallow mantle beneath transform faults⁴². Together with the hiatus in eruption of Pacific lavas across the AAD bounding transform, these observations constrain mantle migration to a narrow, relatively shallow region, directly beneath the spreading axis of the Southeast Indian Ridge. □

Methods

The admittance between free-air gravity and bathymetry can be approximated using: $Z(k) = g(k)/b(k) = 2\pi G \Delta \rho \exp(-kd) + \dots$, where Z is the admittance, k is the wavenumber, $g(k)$ and $b(k)$ are the spectra of the free-air gravity and bathymetry respectively, G is the gravitational constant, $\Delta \rho$ is the average sea-floor density contrast ($\rho_{\text{seafloor}} - \rho_{\text{water}}$), and d is the average depth to the sea floor along the profile. For wavelengths between ~10 and ~100 km, the natural logarithm of the admittance function is a linear function of wavenumber (ref. 43): $\ln(Z) = \ln(2\pi G \Delta \rho) - kd$. We use this spectral relationship between free-air gravity and bathymetry data⁴³ along each axis-crossing profile to compute the average sea-floor rock density by least-squares regression for wavelengths between 10 and 100 km, weighted by the errors in spectral estimation and transfer function reliability⁴⁴. Large errors for individual profiles require, however, that any interpretation be based on regional averages over a number of profiles. The depths calculated by this method are in reasonable agreement with actual depths (Fig. 3), confirming that this approach is sufficiently accurate to resolve the chief differences in sea-floor lithology.

Received 24 July 1997; accepted 11 May 1998.

1. Weissel, J. K. & Hayes, D. E. The Australian–Antarctic Discordance; new results and implications. *J. Geophys. Res.* **79**, 2579–2587 (1974).
2. Forsyth, D. W., Ehrenbard, R. L. & Chapin, S. Anomalous upper mantle beneath the Australian–Antarctic Discordance. *Earth Planet. Sci. Lett.* **84**, 471–478 (1987).
3. Marks, K. M., Vogt, P. R. & Hall, S. A. Residual depth anomalies and the origin of the Australian–Antarctic Discordance. *J. Geophys. Res.* **95**, 17325–17337 (1990).
4. West, B. P., Sempéré, J.-C., Pyle, D. G., Phipps Morgan, J. & Christie, D. M. Evidence for variable upper mantle temperature and crustal thickness in and near the Australian–Antarctic Discordance. *Earth Planet. Sci. Lett.* **128**, 135–153 (1994).
5. Palmer, J., Sempéré, J.-C., Christie, D. M. & Morgan, J. P. Morphology and tectonics of the Australian–Antarctic Discordance between 123° E and 128° E. *Mar. Geophys. Res.* **15**, 121–152 (1993).
6. Sempéré, J.-C., Palmer, J., Christie, D., Morgan, J. P. & Shor, A. Australian–Antarctic Discordance. *Geology* **19**, 429–432 (1991).

7. Sempéré, J.-C., West, B. P. & Géli, L. in *Tectonic, Magmatic, Hydrothermal and Biological Segmentation of Mid-ocean Ridges* (eds MacLeod, C. J., Tyler, P. A. & Walker, C. A.) 1–5 (Geol. Soc., London, 1996).
8. West, B. P. Mantle flow and crustal accretion in and near the Australian Antarctic Discordance. *Ph.D. thesis*. University of Washington, Seattle. (1997).
9. Klein, E. M., Langmuir, C. H., Zindler, A., Staudigel, H. & Hamelin, B. Isotope evidence of a mantle convection boundary at the Australian-Antarctic Discordance. *Nature* **333**, 623–629 (1988).
10. Pyle, D. G., Christie, D. M. & Mahoney, J. J. Resolving an isotope boundary within the Australian-Antarctic Discordance. *Earth Planet. Sci. Lett.* **112**, 161–178 (1992).
11. Hey, R. N., Sinton, J. M. & Duennebie, F. K. in *The Geology of North America: The Eastern Pacific Ocean and Hawaii* (eds Winterer, E. L., Hussong, D. M. & Decker, R. W.) 161–176 (Geol. Soc. Am., Denver, CO, 1989).
12. Blackman, D. K., Cann, J. R. & Janssen, B. The structure of oceanic core complexes: new observations from the Mid-Atlantic Ridge at Atlantis transform fault. *Eos* **78**, F663 (1997).
13. Tucholke, B. E., Lin, J. & Kleinrock, M. C. Relation between rift-axis magmatism, frequency of detachment faulting and development of inside-corner highs in slow-spreading crust. *Eos* **78**, F663 (1997).
14. Pyle, D. G. Geochemistry of mid-ocean ridge basalt within and surrounding the Australian Antarctic Discordance. *Ph.D. thesis*. (Oregon State University, Corvallis. (1994)).
15. Cann, J. R. *et al.* Corrugated slip surfaces formed at ridge-transform intersections on the Mid-Atlantic Ridge. *Nature* **385**, 329–332 (1997).
16. Cannat, M. Emplacement of mantle rocks in the seafloor at mid-ocean ridges. *J. Geophys. Res.* **98**, 4163–4172 (1993).
17. Cannat, M. How thick is the magmatic crust at slow spreading oceanic ridges? *J. Geophys. Res.* **101**, 2847–2857 (1996).
18. Karson, J. A. & Elthon, D. Evidence for variations in magma production along oceanic spreading centers: a critical appraisal. *Geology* **15**, 127–131 (1987).
19. Bonatti, E. Anomalous opening of the Equatorial Atlantic. *Earth Planet. Sci. Lett.* **143**, 147–160 (1996).
20. Tucholke, B. E. & Lin, J. A geological model for the structure of ridge segments in slow-spreading ocean crust. *J. Geophys. Res.* **99**, 11937–11958 (1994).
21. Madsen, J. A., Grindlay, N. R., Rommeveaux, C. & Sclater, J. Morphologic characteristics of ridge segments along the Southwest Indian Ridge, 15–35 degrees E. *Eos* **77**, F671 (1996).
22. Grindlay, N. R., Madsen, J. A., Rommeveaux, C. & Sclater, J. Segmentation of the Southwest Indian Ridge, 15–25° E. *Eos* **77**, F672 (1996).
23. Klein, E. M. & Karsten, J. L. Ocean-ridge basalts with convergent-margin geochemical affinities from the Chile Ridge. *Nature* **374**, 52–57 (1995).
24. Dick, H. J. B. in *Magmatism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 71–105 (Geol. Soc., London, 1989).
25. Schilling, J.-G. Iceland mantle plume: Geochemical evidence along Reykjanes Ridge. *Nature* **242**, 565–571 (1973).
26. Schilling, J.-G., Kingsley, R. H. & Devine, J. D. Galapagos hot spot-spreading center system 1. spatial petrological and geochemical variations (83° W–101° W). *J. Geophys. Res.* **87**, 5593–5610 (1982).
27. Schilling, J.-G. Upper mantle heterogeneities and dynamics. *Nature* **314**, 62–67 (1985).
28. Mahoney, J. J. *et al.* Isotope and tracer element characteristics of a super-fast spreading ridge: East Pacific Rise, 13–23° S. *Earth Planet. Sci. Lett.* **121**, 173–193 (1993).
29. Mahoney, J. J. *et al.* Isotopic and geochemical provinces of the western Indian Ocean spreading centers. *J. Geophys. Res.* **94**, 4033–4052 (1989).
30. Pyle, D. G., Christie, D. M., Mahoney, J. J. & Duncan, R. A. Geochemistry and geochronology of ancient southeast Indian and southwest Pacific seafloor. *J. Geophys. Res.* **100**, 22261–22282 (1995).
31. Lanyon, R., Crawford, A. J. & Eggins, S. M. Westward migration of Pacific Ocean upper mantle into the Southern Ocean region between Australia and Antarctica. *Geology* **23**, 511–514 (1995).
32. West, B. P., Wilcock, W. S. D., Sempéré, J.-C. & Géli, L. Three-dimensional structure of asthenospheric flow beneath the Southeast Indian Ridge. *J. Geophys. Res.* **102**, 7783–7802 (1997).
33. Anderson, R. N., Spariosu, D. J., Weissel, J. K. & Hayes, D. E. The interrelation between variations in magnetic anomaly amplitudes and basalt magnetization and chemistry along the Southeast Indian Ridge. *J. Geophys. Res.* **85**, 3883–3898 (1980).
34. Hayes, D. E. & West, J. K. Depth anomalies in the Southeast Indian Ocean (abstr.). *Geol. Soc. Am. Abstr. Programs* **6**, 784 (1974).
35. Gurnis, M., Müller, R. D. & Moresi, L. Cretaceous vertical motion of Australia and the Australian-Antarctic Discordance. *Science* **279**, 1499–1504 (1998).
36. Anderson, R. N., Spariosu, D. J., Weissel, J. K. & Hayes, D. E. The interrelation between variations in magnetic anomaly amplitudes and basal magnetization and chemistry along the Southeast Indian Ridge. *J. Geophys. Res.* **85**, 3883–3898 (1980).
37. Klein, E. M., Langmuir, C. H. & Staudigel, H. Geochemistry of basalts from the Southeast Indian Ridge, 115° E–138° E. *J. Geophys. Res.* **96**, 2089–2107 (1991).
38. Alvarez, W. Geological evidence for the geographical pattern of mantle return flow and the driving mechanism of plate tectonics. *J. Geophys. Res.* **87**, 6697–6710 (1982).
39. Alvarez, W. Geologic evidence for the plate driving mechanism: the continental undertow hypothesis and the Australian-Antarctic Discordance. *Tectonics* **9**, 1213–1220 (1990).
40. Crawford, A. J., Briquieu, L., Laporte, C. & Hasenaka, T. *Coexistence of Indian and Pacific Oceanic Upper Mantle Reservoirs Beneath the Central New Hebrides Island Arc* 199–217 (*AGU Monogr.* 88, Am. Geophys. Union, Washington, DC, 1995).
41. Hergt, J. M. & Hawkesworth, C. J. Pb, Sr and Nd isotopic evolution of the Lau Basin: Implications for mantle dynamics during backarc opening. *Proc. ODP* **135**, 505–517 (1994).
42. West, B. P., Christie, D. M. Diversion of along-axis asthenospheric flow beneath migrating ridge-transform-ridge intersections. *Eos* **78**, F673 (1997).
43. McKenzie, D. P. & Bowin, C. The relationship between bathymetry and gravity in the Atlantic Ocean. *J. Geophys. Res.* **81**, 1903–1915 (1976).
44. Wessel, P. & Smith, W. H. F. Free software helps map and display data. *Eos* **72**, 445–446 (1991).

Acknowledgements. This work was supported by the US NSF.

Correspondence and requests for materials show be addressed to D.M.C. (e-mail: dchristie@oce.orst.edu).

Crystal structure of the spliceosomal U2B''–U2A' protein complex bound to a fragment of U2 small nuclear RNA

Stephen R. Price*, Philip R. Evans & Kiyoshi Nagai

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

We have determined the crystal structure at 2.4 Å resolution of a ternary complex between the spliceosomal U2B''/U2A' protein complex and hairpin-loop IV of U2 small nuclear RNA. Unlike its close homologue the U1A protein, U2B'' binds to its cognate RNA only in the presence of U2A', which contains leucine-rich repeats in its sequence. The concave surface of a parallel β-sheet within the leucine-rich-repeat region of U2A' interacts with the ribonucleoprotein domain of U2B'' on the surface opposite its RNA-binding surface. The basic carboxy-terminal region of U2A' interacts with the RNA stem. The crystal structure reveals how protein–protein interaction regulates RNA-binding specificity, and how replacing only a few key residues allows the U2B'' and U1A proteins to discriminate between their cognate RNA hairpins by forming alternative networks of interactions.

The ribonucleoprotein (RNP) motif, also known as RRM (for RNA-recognition motif) or RNP-CS RBD (RNP consensus-sequence-type RNA-binding domain), is one of the most common protein sequence motifs^{1–5}. The RNP domain, found either on its own or in multiple copies^{6,7}, function as an RNA-binding module in many proteins involved in RNA processing and transport⁸. Two highly conserved short sequences, the RNP1 and RNP2 motifs which are embedded in a weakly conserved 80-amino-acid motif, characterize the RNP domain, which folds into a compact structure containing a four-stranded antiparallel β-sheet flanked on one side by two α-helices with the RNP1 and RNP2 motifs in the two middle β-strands⁹. Proteins containing the RNP domain recognize specific RNA targets with various secondary structures, including hairpin loops, internal loops and single strands⁸. How did the RNP motif evolve to recognize different targets with high specificity?

The crystal structure of a 98-residue fragment of the spliceosomal U1A protein in complex with U1 snRNA hairpin II determined at 1.92 Å resolution revealed important principles of RNA recognition by the RNP motif¹⁰. U1 snRNA hairpin II contains a 10-nucleotide loop, with the AUUGCACUCC sequence closed by a C-G base pair. The RNA loop binds to the surface of the protein β-sheet as an open structure, the polypeptide joining the β2 and β3 strands protruding through the RNA loop. The first seven loop nucleotides, AUUGCAC, and the loop-closing C-G base pair form an intricate hydrogen-bond network with protein side chains and main chain¹⁰. These nucleotides show stacking interactions with either an adjacent RNA base or a protein side chain, or both, which restrict the orientation of the RNA bases, hence stabilizing the hydrogen-bonding network with the protein. In addition, the RNA loop and stem touch the most basic region of the protein, showing the importance of electrostatic interactions.

The U2B'' protein¹¹, a component of the spliceosomal U2 small nuclear (sn) RNP, is closely related to the U1A protein¹²; both contain two RNP motifs at their amino and carboxy termini, linked by a protease-sensitive polypeptide. The U2B'' protein binds hairpin IV of U2 snRNA only when it is complexed with the U2A' protein^{13–16}, which is a member of the leucine-rich repeat (LRR) protein family^{17–19}. A fragment of U1A containing the first 98 residues is both necessary and sufficient for its specific binding to hairpin II of U1 snRNA²⁰. Similarly, the corresponding N-terminal

fragment of U2B'' protein is fully active in forming a complex with U2LRR protein and the resulting complex shows high affinity and specificity for hairpin IV of U2 snRNA^{13,14} (to avoid confusion, we shall refer to U2A' protein as U2LRR). There is no crosstalk between these two systems; neither the U1A nor the U2B''/U2LRR complex bind to non-cognate RNA hairpins¹⁴. There are 25 amino-acid replacements between the U1A and U2B'' proteins within this domain^{11,12}. RNA-binding studies of chimaeras of U1A and U2B'' proteins have identified amino-acid residues important for the discrimination between the two RNA hairpins and for the interaction with the U2LRR protein^{13,14,21,22}.

The LRR motif is found in multiple copies in many proteins known to form complexes with other proteins. Typically, each LRR motif consists of 22–29 amino acids with the consensus sequence of XLXLXLLXXN where L, N and X represent leucine, asparagine and any amino acid, respectively^{17–19}. Porcine liver ribonuclease inhibitor consists entirely of LRRs; it is the only protein in this family of known structure, both on its own²³ and as a complex with RNase A²⁴. Each of its fifteen LRR motifs forms a β-strand packed against an α-helix, with the leucine side chains at conserved positions making a hydrophobic core; their assembly forms a horseshoe-shaped solenoid with its inner circumference made of a parallel β-sheet and its outer circumference of α-helices²³. U2LRR protein contains only five LRR motifs which are flanked by unrelated sequences^{14,16}.

We have solved the crystal structure of a ternary complex between U2B'' and U2LRR proteins with U2 snRNA hairpin IV. A comparison of the crystal structures of the ternary complex and the complex between the U1A protein and its cognate RNA hairpin reveals the stereochemical basis of the discrimination between the two highly similar RNA targets, and has given insights into the mechanism of RNA–protein recognition and LRR–protein recognition.

Overall structure of the complex

Fragments of human U2B'' (N-terminal 96 residues)¹¹ and U2A'(U2LRR)¹⁵ (residues 1–176) proteins were co-expressed in *Escherichia coli*, purified and crystallized with a 24-nucleotide RNA (rCCUGGUAUUGCAGUACCUCAGGU) (Fig. 1). The crystals belong to the orthorhombic space group *P*₂₁₂₁, with unit cell edges of *a* = 98.4, *b* = 128.2, *c* = 66.7 Å. The overall structure of one of the two ternary complexes in the asymmetric unit is shown in stereo in Fig. 2a.

In U2LRR, the LRR region resides between residues 22 and 146 (Fig. 1b) and forms a solenoid similar to, but more irregular than,

* Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, 701 West 168th Street, New York, New York 10032, USA.

that of the ribonuclease inhibitor^{14–16} (Fig. 2a, b). The parallel β -sheet within the LRR region is canonical¹⁷ but the chains forming the outer circumference form either a 3_{10} helix, a short α -helix, or more extended, irregular structures^{17,18}, and $\beta 6$ and $\beta 7$ are joined by two helices. The N-terminal flanking region forms an α -helix and a long polypeptide loop which folds back and forms hydrogen bonds with the β -strand ($\beta 2$) of the first LRR motif. The U2LRR protein forms five complete LRRs and terminates in the middle of the sixth β -strand ($\beta 7$), which is followed by two α -helices (Fig. 1b). Helix A of U2B' protein fits into the concave surface of the parallel β -sheet of the LRR region and the N- and C-terminal arms of U2LRR protein embrace the U2B' RNP domain (Fig. 2b). In contrast, the U1A protein does not form a complex with the U2LRR protein; however, two amino-acid replacements, Asp 24 $\rightarrow$ Glu and Lys 28 $\rightarrow$ Arg, allow it to form a stable complex²¹. The guanidinium side chain of Arg 28 of U2B' protein protrudes to form hydrogen bonds with Thr 69 and Glu 92 of the U2LRR protein. The corresponding Lys 28 of U1A protein cannot form such an extensive interaction. Glu 24 of U2B' protein extends towards Arg 20 of U2LRR protein, although no salt bridge is formed between these two residues.

The interface between RNA and protein

The U2 snRNA hairpin IV binds to U2B' in roughly the same orientation as hairpin II of U1 snRNA binds to U1A protein¹⁰, but their interactions differ in detail (Fig. 2c, d). Hairpin II of U1 snRNA and hairpin IV of U2 snRNA have the first six loop nucleotides in common (Fig. 1c, d). Residues that differ between U2B' (yellow in Fig. 2c) and U1A (red in Fig. 2d) proteins form two clusters at or close to the RNA contact. In one region, six consecutive amino-acid residues (residues 44–49) at the end of the $\beta 2$ strand differ between the U1A and U2B' proteins; these residues in the U2B' protein are in contact with the 3' half of the RNA loop. Leu 17 and Glu 19 of U1A protein near the loop-closing base pair are replaced by Met and Asp in the U2B' protein. These two regions are important for the discrimination between the two RNA hairpins^{13,21,22,25}. The recognition elements of hairpin II of U1 snRNA and hairpin IV of U2 snRNA can be divided into four regions. We shall discuss each in turn.

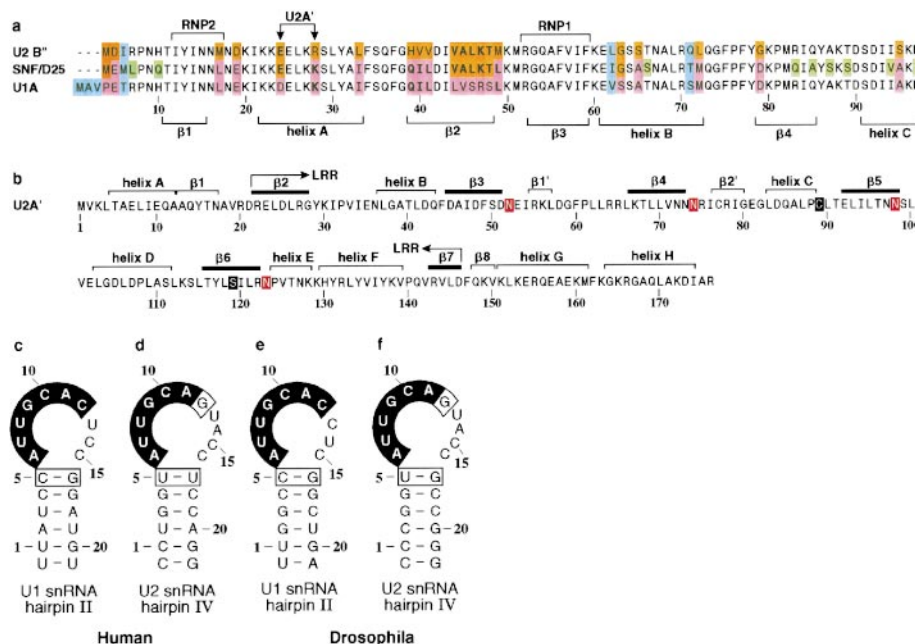
Conserved interactions with U8, G9, C10 and A11. The RNA–protein contacts involving U8, G9, C10 and A11 are largely conserved between the two complexes¹⁰. Phe 56, A11, C/G12 and the side chain of Asp 92 form a continuous stack and C10 stacks onto Tyr 13 in both U1A/RNA (Fig. 5e of ref. 10) and U2B'/U2LRR/

RNA complexes (Fig. 3a). A hydrogen bond between the 2-amino group of G12 and the phosphate group stabilizes G12 in the *syn* conformation²⁶, allowing the larger purine base of G12 to fit into the space occupied by C12 in the U1A complex. The O6 of G12 forms a hydrogen bond with the main-chain amide of Asp 92 and the ϵ -amino group of Lys 50 forms a salt bridge with the phosphate of A11. **Interactions of the 3' half of the RNA loop.** The last three loop nucleotides (U13, C14 and C15) of U1 snRNA hairpin II are poorly ordered in the U1A–RNA complex¹⁰, whereas in the U2 ternary complex the last four loop nucleotides (U13, A14, C15, C16) form a rigid structure resembling a stepladder (Fig. 3b). This is stabilized by stacking interaction between A14, C15 and C16 and a network of hydrogen bonds involving the phosphates of C15 and C16 and the base of U13. The bases of C16 and U13 pack against the γ -methyl group of Thr 48 (Fig. 3d) and the side chain of Leu 46 (omitted in Fig. 3b for clarity), respectively. Both Leu 46 and Thr 48 within the six-amino-acid segment (residues 44–49) are replaced by Ser in U1A, reducing the affinity for hairpin IV of U2 snRNA by destabilizing the stepladder structure. The replacement of C12 with G alters the RNA backbone conformation, allowing U13–C16 to form the stepladder structure. This structure is important in positioning the loop-closing U5–U17 base pair differently from the corresponding C5–G16 base pair in the U1A–RNA complex (Fig. 3c, d).

Interaction of the loop-closing base pair with the protein. The 11-nucleotide loop of the U2 snRNA hairpin IV is closed by an asymmetric non-Watson–Crick U–U base pair²⁶ (Fig. 3d). The backbone atoms of the U–U base pair are much closer together (C_1' – C_1' distance, 8.8 Å) than those of the loop-closing C–G base pair in the U1A–RNA complex (10.5 Å). This causes the first base of the loop (A6) to stack against U17 on the opposite strand rather than onto the adjacent loop-closing base, as seen in the U1A–RNA complex. The side chain of Lys 20 points towards the O4 atoms of U17 and U5 and the O6 of the adjacent G4.

In the U1A complex, Arg 52 hydrogen-bonds with O6 and N7 of G16 as well as N1 of A6 (Fig. 3c)¹⁰, whereas Arg 52 of the U2B' protein forms salt bridges with the phosphate group of U17 and the carboxyl group of Asp 19 (Fig. 3d, f). In the U1A protein, the corresponding Glu 19 hydrogen bonds with the 3-imino group of U7 and the 2-amino group of G9 (Fig. 3c, e). Two conservative changes, Glu 19 $\rightarrow$ Asp and Met 17 $\rightarrow$ Leu, and the replacement of the loop-closing C–G base pair with a U–U base pair cause a rearrangement of the hydrogen-bonding network involving conserved residues in the two complexes. The different positioning of

Figure 1 Amino-acid sequences of the U2B' and U2A' (U2LRR) proteins and their RNA binding site. **a**, Amino-acid sequence alignment of the RNA binding fragments of the spliceosomal U2B', *Drosophila* SNF/D25 and U1A proteins. Residues of U2B' and SNF/D25 proteins are numbered based on U1A protein numbering. Pink, U1A specific; yellow, U2B' specific; green, unique to *Drosophila* SNF/D25 protein; blue, variable. **b**, Amino-acid sequence of U2A' (U2LRR) protein with secondary-structure elements. $\beta 2$ – $\beta 7$, Segments containing the β -strand and part of the β -turn, which form a sheet together with the $\beta 1$ and $\beta 8$ strands. $\beta 1'$ and $\beta 2'$ from a short sheet. The conserved asparagines are highlighted in red and the two residues (Cys 89 and Ser 119) that were mutated to make a heavy-atom derivative are also highlighted. **c**, Human U1 snRNA hairpin II. **d**, Human U2 snRNA hairpin IV. **e**, *Drosophila* U1 snRNA hairpin II. **f**, *Drosophila* U2 snRNA hairpin IV. Nucleotides are numbered so that the conserved nucleotides in the loop carry the same numbers.



the loop-closing base pairs allows the RNA stem to extend in different directions in the U1A and U2B''/U2LRR complexes (Fig. 2c, d).

Interactions between the RNA stem and the proteins. The double-stranded stem of the U2 snRNA hairpin IV interacts with both U2B'' and U2LRR proteins, as predicted¹⁴. The surface electrostatic potential of the U2B''/U2LRR protein complex shown in Fig. 4a indicates that the electrostatic interaction between the U2LRR protein and the RNA stem is important for the formation of the ternary complex. The ϵ -amino groups of Lys 22 of U2B'' protein and Lys 151 of the U2LRR protein form salt bridges with the phosphate groups of C1 and G3 in the RNA stem, respectively. The Asp 24 $\rightarrow$ Glu and Lys 28 $\rightarrow$ Arg mutations enable the U1A protein to form a complex with the U2LRR protein, which substantially

weakened the binding of U1 snRNA hairpin II²¹. Figure 4c shows a modelled structure of a ternary complex generated by docking the U1A-hairpin II RNA complex with U2LRR protein. The stem of U1 snRNA hairpin II is unable to interact with the U2LRR protein in the same way as in the U2 complex. The electrostatic attraction between the RNA stem and the basic U2LRR protein would distort the interaction of the protein with the loop-closing base pair and loop nucleotides, weakening their binding.

Discrimination of four recognition elements

The specific binding of the U1A and U2B'' proteins to their cognate RNAs is crucial for the correct assembly and integrity of the U1 and U2 snRNPs. The U₈G₉C₁₀A₁₁ sequence common to both the U1 snRNA

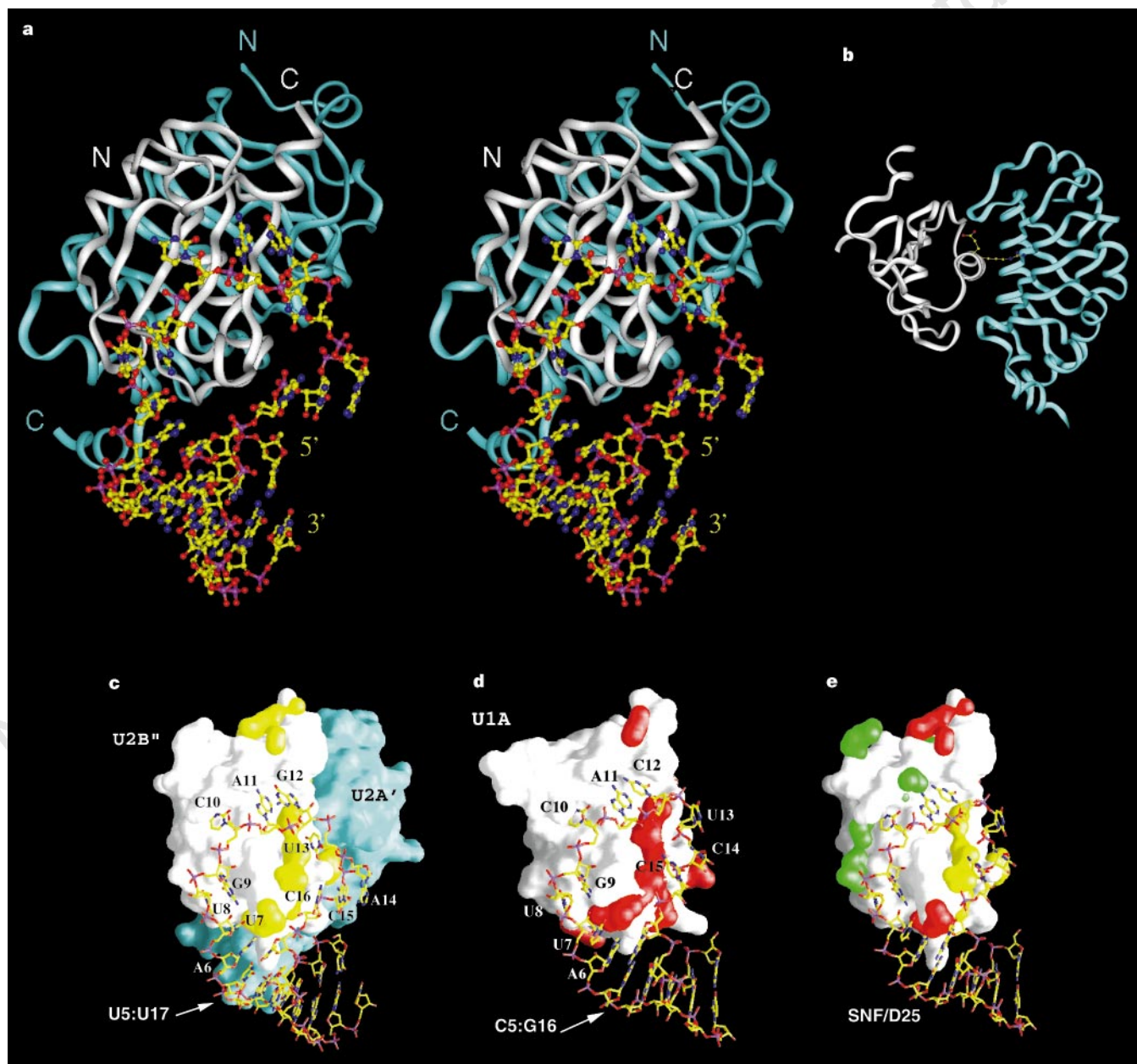


Figure 2 Overall structure of the U2B''-U2A' (U2LRR) protein complex bound to hairpin IV of U2 snRNA. **a**, A stereo pair showing the overall structure of the ternary complex (ABQ complex). White ribbon, U2B'' protein; blue ribbon, U2A' (U2LRR) protein; ball-and-stick, U2 snRNA hairpin IV. **b**, The interaction between U2A' (U2LRR) and U2B'' proteins. The two residues, Glu24 and Arg28, that are important for this interaction are included. **c**, A surface representation of the ternary complex between U2B''/U2A' proteins and U2 snRNA hairpin IV. Blue, U2A' (U2LRR) protein; white, U2B'' protein, with amino-acid residues specific to

U2B'' protein highlighted in yellow; **d**, a surface representation⁴⁹ of a complex between U1A protein and U1 snRNA hairpin II (ref. 10). Amino-acid residues specific to U1A protein are highlighted in red; **e**, a model of a complex between *Drosophila* SNF/D25 protein and U1 snRNA hairpin II. The structure of the protein is based on the U2B'' protein but amino-acid side chains are colour-coded as in Fig. 1a. Red, U1A specific; yellow, U2B'' specific; green, unique to *Drosophila* SNF/D25; white, conserved residues.

hairpin II and the U2 snRNA hairpin IV form nearly identical interactions with residues conserved between the U1A and U2B' proteins. What are the additional recognition elements recognized by U1A protein? An *in vitro* RNA selection experiment using a pool of RNA hairpins with 10 randomized loop nucleotides showed that the AUUGCAC sequence in the first seven positions is strongly selected²⁷, whereas there is no preference in the last three loop positions. Indeed, the last three poorly ordered loop nucleotides can be replaced by an ethylene glycol linker without significantly altering the binding affinity²⁸. Hence, biochemical data are consistent with the crystal structure, showing that the AUUGCAC sequence and the loop-closing C-G basepair are the only base-specific elements recognized by the U1A protein. These recognition elements are conserved in the U1A protein-binding sites within the 3' untranslated region of its own mRNA^{29–32}.

The RNA contacts are much more extensive in the U2B'/U2LRR complex than in the U1A complex. In particular, the last four loop nucleotides and the RNA stem of U2 snRNA make extensive interactions not present in the U1 snRNA counterpart¹⁰. If the isolated U2B' protein could form all these contacts, it should bind the U2 snRNA hairpin IV at least as strongly as does U1A protein to U1 snRNA hairpin II. The fact that the U2B' protein absolutely requires the U2LRR protein for RNA binding¹³ shows that the binding of the four recognition elements of the U2 snRNA hairpin IV to the U2B'/U2LRR protein complex is highly cooperative. The C-terminal region of the U2LRR protein is strongly basic (Fig. 4a) and the hydrogen bonds and salt bridges between the RNA stem and the U2B'/U2LRR protein complex are crucial in positioning the RNA stem (Fig. 4b). These interactions allow the formation of the intricate hydrogen-bond network involving A6, U7 and the loop-closing nucleotides, and, in turn, allow the formation of the 'stepladder' structure by the last four loop nucleotides. These nucleotides are packed

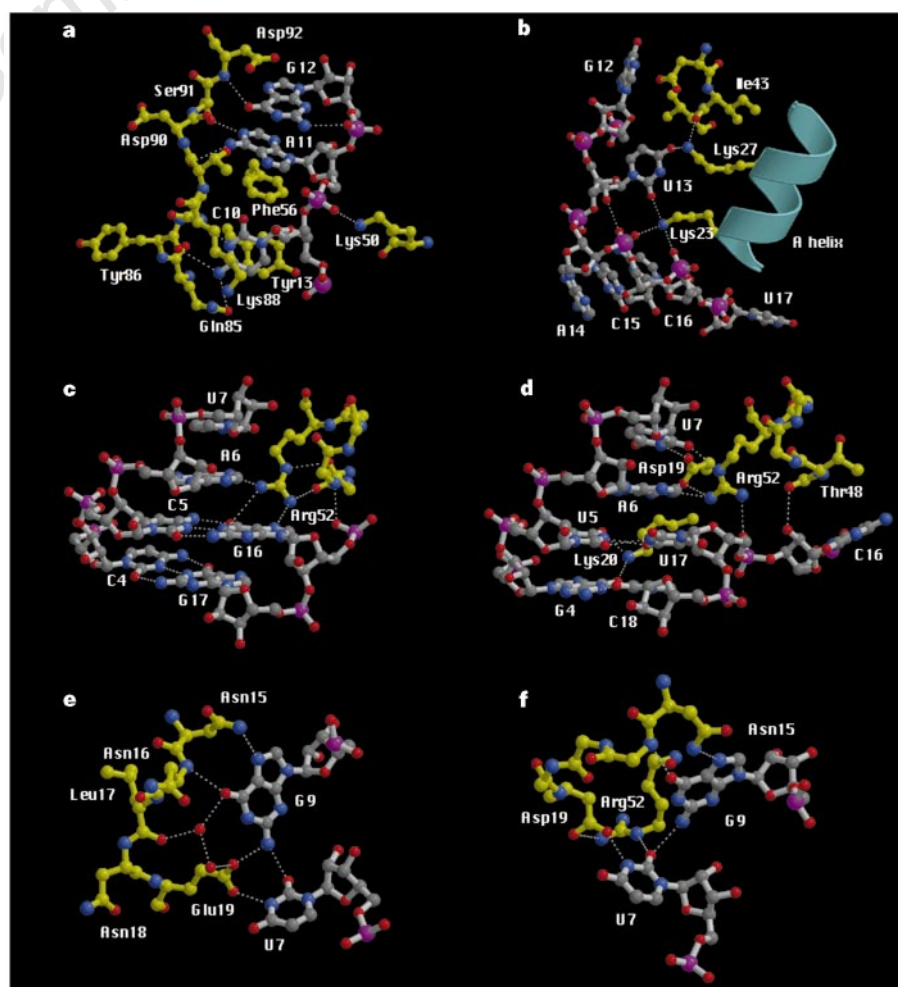
against the six-amino-acid segment of the U2B' protein, VALKTM (residue 44–49) and its replacement with LVSRSL, found in the U1A protein, substantially reduces the binding of U2 snRNA^{13,21,25,33}, owing to the loss of van der Waals contacts. This would in turn affect the hydrogen-bond network involving the loop-closing nucleotides and would change the orientation of the RNA stem.

Drosophila SNF/D25 protein

Drosophila contains neither U1A nor U2B' proteins: a single protein, SNF/D25, serves the function of both (Fig. 1a). The *Drosophila* SNF/D25 protein binds to *Drosophila* U1 snRNA (Fig. 1e) *in vitro*³⁴. It has been shown³⁵ that SNF/D25 is a component of both U1 and U2 snRNPs and *in vitro* it binds hairpin IV of *Drosophila* U2 snRNA (Fig. 1f) in the presence of human U2LRR protein or *Drosophila* nuclear extract³⁵. *Drosophila* U2LRR protein has not been cloned, but a homologue of the U2LRR protein is likely to be present in *Drosophila* nuclear extract. Figure 2e shows a model of the SNF/D25 protein in complex with the U1 snRNA, in which the amino-acid side chains are coloured as in Fig. 1a. The residues in contact with the 3' half of the loop are of the U2B' protein type (yellow), whereas those that interact with the loop-closing base pair are of the U1A protein type (red). At the RNA interface, there are three residues unique to the SNF/D25 protein (green): Gln 85 → Ala, Ala 87 → Ser and Thr 89 → Ser. The loss of a hydrogen bond between Gln 85 and the 4-amino group of C10 found in both human U1A and U2B'/U2LRR complexes is compensated for by hydrogen bonds between Ser 87 and the 2-carbonyl group and/or 4-nitrogen of C10 (Fig. 1 3a). The hydrogen bond made by Thr 89 would be maintained by the Thr 89 → Ser substitution.

The *Drosophila* U1 snRNA hairpin II shares the sequence AUUGCAC and the same loop-closing C-G base pair with the human

Figure 3 Comparison of the RNA-protein interfaces between the U2B'/U2LRR(U2A')/RNA complex and the U1A protein-RNA complex¹⁰. **a**, Stacking and hydrogen-bonding interactions of C10, A11 and G12 in the U2B'/U2LRR/RNA complex; **b**, the stepladder structure with three stacked nucleotide bases, A14, C15 and C16; **c**, the hydrogen-bonding interactions of the loop-closing C5-G16 base pair in the U1A-RNA complex¹⁰; **d**, the intricate hydrogen-bonding interactions of the loop-closing U5-U17 base pair in the U2B'/U2A'/RNA complex. **e**, The hydrogen-bonding network of U7, G9 and Glu 19 in the U1A-RNA complex¹⁰; **f**, the hydrogen-bonding network of U7, G9, Asp 19 and Arg 52 in the U2B'/U2LRR/RNA complex. These figures are drawn using Bobscript⁶⁰.



counterpart. The *Drosophila* SNF/D25 protein contains Leu 17, Glu 19 and Leu 49, which are important for the binding of human U1A protein to U1 snRNA, and therefore the RNA–protein contacts shown in Fig. 3c, d are likely to be conserved in the *Drosophila* SNF/D25–U1 snRNA complex. Arg 52 would form hydrogen bonds with G16, and Glu 19 with U7 and G9 as in the human U1A complex. The replacement of the LVSRSL sequence (residues 44–49) in U1A with the U2B'' type (VALKTM) reduces the affinity for the U1 snRNA hairpin II^{13,21,25}, but Leu 49, which is the most important of these amino acids for binding of U1 snRNA³⁶, is retained in SNF/D25. This explains why SNF/D25 binds *Drosophila* U1 snRNA hairpin II.

The *Drosophila* U2 snRNA hairpin IV is closed by a U–G wobble base pair^{26,37} and contains the same last five loop nucleotides, GUACC, as the human U2 snRNA hairpin IV. The UACC sequence could readily form the stepladder structure with three stacked bases and is likely to be packed against the VALKT sequence conserved between the human U2B'' and SNF/D25 proteins (Fig. 3b). The loop-closing U–G wobble base pair (Fig. 1f) projects three hydrogen bond acceptors into the major groove, and Arg 52 would form hydrogen bonds either with O6 and N7 of G17 or with O6 of G17 and O4 of U5. Glu 19 would form hydrogen bonds with U7 and G9, as in the U1A protein–hairpin complex (Fig. 3e). The replacement of the loop-closing C–G base pair with a U–G wobble base pair changes the position of the 5' strand of the RNA stem with respect to G17 and hence would alter the orientation of the RNA stem, allowing the RNA stem to interact with the C-terminal region of the U2LRR protein as in the human U2B''/U2LRR protein complex. SNF/D25 has only one of the residues (Glu 24) important for the binding to the U2LRR protein²¹ and thus it would bind to the U2LRR less tightly than the U2B'' protein, allowing the formation of both the U1 snRNA complex and the ternary complex with the U2LRR protein and U2 snRNA. U2LRR thus acts as a true modulator of RNA-binding specificity in this system, allowing the SNF/D25 protein to present alternative RNA specificity determinants.

Evolution of specificity

The U1A and U2B'' proteins achieve highly specific discrimination

Figure 4 Interaction between U2A' (U2 LRR) protein and the RNA stem in the U2B''/U2LRR/RNA and hypothetical U1A/U2LRR/RNA complexes. **a**, A surface representation⁴⁹ of the U2B''/U2A'(U2LRR)/RNA complex with electrostatic potential. Blue, positive; red, negative. **b**, A surface representation of the U2B''/U2LRR/RNA complex showing the interaction between U2LRR protein and the RNA stem. **c**, A model of a complex between U2LRR protein and the U1A–RNA complex¹⁰ generated by computer graphics docking. The surface is colour-coded as in Fig. 2c, d.

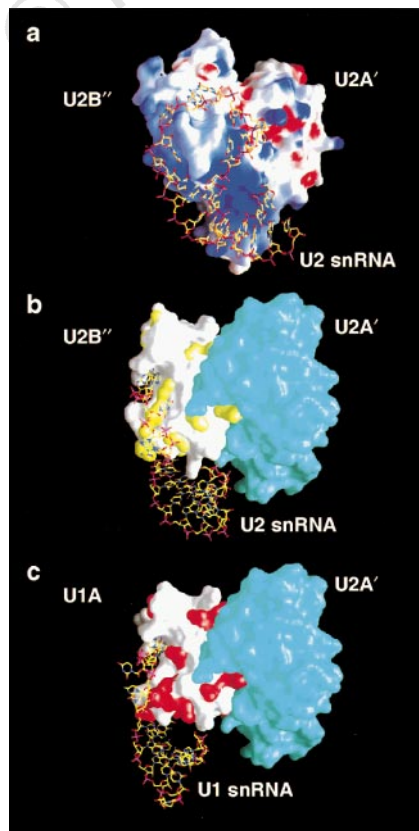


Table 1 X-ray structure determination

Spacegroup <i>P</i> 2 ₁ 2 ₁ 2		Cell: <i>a</i> = 98.4, <i>b</i> = 128.2, <i>c</i> = 66.7 Å	
U2B''–RNA complex	Native I	Native II	Derivative
Resolution range	76.7–2.38 Å	79–2.7 Å	39.1–2.8 Å
Wavelength	1.000 Å	0.87 Å	0.990 Å
Number of unique reflections	31,458	23,742	20,831
Mean redundancy	4.6	4.3	4.3
Completeness (outer shell)	95.6% (74%)	99.6% (99.2%)	99.2% (99.5%)
Overall <i>R</i> _{sym} * (outer shell)	0.061 (0.24)	0.042 (0.11)	0.039 (0.089)
<i>R</i> _{diff} (native II vs derivative)†			0.247
Phasing			
Number of sites			2
Phasing power‡ (anomalous)	Acentric		1.1 (2.1)
	Centric		0.8
Cullis <i>R</i> factor§			0.73
Mean FOM	Overall		0.433
	Acentric		0.425
	Centric		0.490
Solvent flattening			
Number of cycles			10
Resolution range			17.2–2.8 Å
Mean overall FOM			0.786

**R*_{sym} = $\sum_i |I_i - \langle I \rangle| / \sum_i I_i$, where *I* is the observed intensity and $\langle I \rangle$ is the average intensity from multiple measurements.

†*R*_{diff} = $\sum_i |F_{PH}| - |F_P| / \sum_i |F_P|$.

‡Phasing power is root-mean-square ($|F_h|/E$) where $|F_h|$ is the heavy-atom structure-factor amplitude and *E* is the residual lack of closure error.

§Cullis *R* factor = $(\langle E \rangle / D_{iso})$, where $D_{iso} = |F_{PH}| - |F_P|$.

||FOM, figure of merit.

between their cognate and non-cognate RNA hairpins by mutually induced fit. Both the protein and RNA components become more ordered upon binding^{9,10,32,38,39}, allowing highly cooperative formation of alternative networks of interactions in the two cognate complexes. As is commonly observed in DNA–protein complexes, most of the hydrogen-bond donors and acceptors at the RNA–protein interface are satisfied in both cognate complexes.

If nucleotides that are different in the two RNAs formed contacts with the amino acids that are different in the two proteins, and contacts between conserved nucleotides and amino acids are maintained in the two cognate complexes, then the formation of non-cognate complexes would be only slightly energetically unfavourable. In such a case, highly specific discrimination between the cognate and non-cognate binding sites could not be achieved. In the U1A and U2B''/U2LRR systems, the two cognate complexes form distinct, intricate networks of interactions that could not be achieved by the non-cognate complexes. Our results show that subtle changes in the protein (for example, the Asp 19 → Glu, Met 49 → Leu substitutions) and the RNA (the replacement of the loop-closing base pair) rearranges the hydrogen-bond networks. Even the conserved amino acids and nucleotides form different contacts in the two complexes. This gives rise to highly specific discrimination between the cognate and non-cognate binding sites. This rearrangement is facilitated by the interaction of U2LRR with the RNA stem, which is only possible by the interaction of U2B'' with U2LRR. In *Drosophila*, this is achieved by the single SNF/D25 protein.

A similar recognition mechanism is used by DNA-binding protein families. For example, oestrogen and glucocorticoid hormone receptors are homologous, recognizing closely related DNA sequences, but achieve highly specific discrimination of their DNA targets by forming distinct networks of interactions, as observed here^{40,41}. Undoubtedly, many RNA and DNA-binding protein families⁴² have evolved from a common ancestor to recognize closely related sites in this highly economical manner. □

Methods

Human U2B'' protein¹¹ (the N-terminal 96-residue fragment containing residues 4–99 (the U1A protein numbering) was co-expressed in *E. coli* with human

U2LRR (U2A') protein¹⁵ lacking the ten C-terminal residues. This complex tends to aggregate in the presence of RNA because of the highly charged C-terminal tail. Domain mapping of U2A' protein with proteases showed that the region around residue 161–180, which is susceptible to chymotrypsin and trypsin, corresponds to a domain boundary. A complex of a fragment of U2B'' protein containing residues 4–99 (Fig. 1a) and a fragment of U2A' protein containing residues 1–176 binds hairpin IV of U2 snRNA specifically and with high affinity without aggregation, and was used for crystallization. A 24-nucleotide RNA (rCCUG-GUAUUGCAGUACCUCCA GGU) was synthesized and purified as described⁴³ and used for crystallization⁴⁴. The ternary complex was crystallized in a sitting drop from 50 mM NaCl, 9 mM MgCl₂, 0.25 mM spermine, 0.25% *n*-octyl-β-D-glucopyranoside, 50 mM Tris-Cl, pH 7.3, 1% PEG 600, by batch method. Details of our crystallization procedure will be published elsewhere (S.R.P. *et al.*, manuscript in preparation). Wild-type crystals suffer a phase transition from a centred lattice to a primitive lattice upon heavy-atom soak. This precluded the calculation of X-ray phases using isomorphous replacement with the wild-type proteins. Crystals of the ternary complex containing a U2A' mutant (Cys 89 → Asp and Ser 119 → Cys) belong to orthorhombic space group *P*₂₁₂₁₂, with unit cell edges of 98.4 × 128.2 × 66.7 Å, and did not undergo a phase transition upon heavy-atom derivatization. These mutations affected neither U2A'–U2B'' affinity nor RNA binding. The crystal was soaked stepwise in buffer containing 5% MPD, 10% MPD, 10% MPD + 5% glycerol, 15% MPD + 7.5% glycerol, and finally 20% MPD + 10% glycerol and flash-cooled to 100 K for data collection. An isomorphous derivative was obtained by first soaking crystals in a mother liquor containing 1 mM methylmercury nitrate, and this single derivative was used to determine the crystal structure. All data were processed using programs incorporated within the CCP4 package⁴⁵ (Table 1). Heavy-atom positions were located manually by Patterson methods; these positions were refined and an initial phase set calculated using the program SHARP⁴⁶. Phases were improved using two-fold symmetry averaging and density modification with the program SOLOMON⁴⁷.

The resulting experimental electron density map was of good quality and unambiguously allowed the tracing of a polypeptide chain for the U2A' protein using the program O. The U2B'' protein was fitted using the coordinates of the U1A protein in the U1A protein–RNA complex and mutated to the U2B'' residues using the graphics interface O⁴⁸. Refinement was carried out to higher resolution using a dataset to 2.38 Å resolution. Several rounds of rebuilding and refinement with the program REFMAC⁴⁷ with non-crystallographic restraints between the two complexes in the asymmetric unit led to a crystallographic *R* factor of 28.2% and a free *R* factor of 32.8% for a model consisting of all the RNA residues (0–23), A-chain (U2A') residues 2–163, C-chain 2–175, B-chain (U2B'') residues 6–99, D chain 6–98, with no water molecules (although evidence for some waters can be seen in the maps). The effective resolution of the model is limited by the high temperature factor of the crystals (*B* factor from Wilson plot, 51 Å²; average *B* for model, 56 Å²), and the relatively poor quality is shown by the rather high *R* factors. However, the RNA–protein interface has relatively low *B* factors and all the structural features discussed here are clear in the experimental electron density maps.

Received 11 February; accepted 12 June 1998.

- Uhlenbeck, O. C., Pardi, A. & Feigon, J. RNA structure comes of age. *Cell* **90**, 833–840 (1997).
- Nagai, K. RNA–protein complexes. *Curr. Opin. Struct. Biol.* **6**, 53–61 (1996).
- Burd, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621 (1994).
- Birney, E., Kumar, S. & Krainer, A. Analysis of the RNA-recognition motif and RS and RGG domains—Conservation in metazoan pre-messenger-RNA splicing factors. *Nucleic Acids Res.* **21**, 5803–5816 (1993).
- Mattaj, J. W. RNA recognition: A family matter? *Cell* **73**, 837–840 (1993).
- Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nature Struct. Biol.* **4**, 215–222 (1997).
- Xu, R.-M., Jokhan, L., Cheng, X., Mayeda, A. & Krainer, A. R. Crystal structure of human UPI, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* **5**, 559–570 (1997).
- Nagai, K., Oubridge, C., Ito, N., Avis, J. & Evans, P. The RNP domain—A sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci.* **20**, 235–240 (1995).
- Nagai, K., Oubridge, C., Jessen, T.-H., Li, J. & Evans, P. R. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515–520 (1990).
- Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438 (1994).
- Habets, W. J. *et al.* Analysis of a cDNA clone expressing a human autoimmune antigen—Full-length sequence of the U2 small nuclear RNA-associated B'' antigen. *Proc. Natl Acad. Sci. USA* **84**, 2421–2425 (1987).
- Sillekens, P. T. G. & Habets, W. J., Beijer, R. P. & van Venrooi, W. J. cDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins.

EMBO J. **6**, 3841–3848 (1987).

- Scherly, D., Boelens, W., Dathan, N. A., van Venrooi, W. J. & Mattaj, J. W. Major determinants of the specificity of interaction between small nuclear ribonucleoprotein-U1A and ribonucleoprotein-U2B'' and their cognate RNAs. *Nature* **345**, 502–506 (1990).
- Boelens, W. *et al.* A weak interaction between the U2A' protein and U2 snRNA helps to stabilize their complex with the U2 B'' protein. *Nucleic Acids Res.* **19**, 455–460 (1990).
- Sillekens, P. T. G., Beijer, R. P., Habets, W. J. & van Venrooi, W. J. Molecular cloning of the cDNA for the human U2 snRNA-specific A' protein. *Nucleic Acids Res.* **17**, 1893–1906 (1989).
- Fresco, L. D., Harper, D. S. & Keene, J. D. Leucine periodicity of U2 small nuclear ribonucleoprotein particle (snRNP) A'-protein is implicated in snRNP assembly via protein–protein interactions. *Mol. Cell. Biol.* **11**, 1578–1589 (1991).
- Kobe, B. & Deisenhofer, J. Proteins with leucine-rich repeats. *Curr. Opin. Struct. Biol.* **5**, 409–416 (1995).
- Kajava, A. V., Vassart, G. & Wodak, S. J. Modeling of the 3-dimensional structure of proteins with the typical leucine-rich repeats. *Structure* **3**, 867–877 (1995).
- Buchanan, S. G. S. & Gay, N. J. Structural and functional diversity in the leucine rich repeat family of proteins. *Prog. Biophys. Molec. Biol.* **65**, 1–44 (1996).
- Scherly, D. *et al.* Identification of the RNA binding segment of human U1A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**, 4163–4170 (1989).
- Scherly, D., Dathan, N. A., Boelens, W., van Venrooi, W. J. & Mattaj, J. W. The U2B'' RNP motif as a site of protein–protein interaction. *EMBO J.* **9**, 3675–3682 (1990).
- Scherly, D., Kambach, C., Boelens, W., van Venrooi, W. J. & Mattaj, J. W. Conserved amino acid residues within and outside of the N-terminal ribonucleoprotein motif of U1A small nuclear ribonucleoprotein involved in U1 RNA binding. *J. Mol. Biol.* **219**, 577–584 (1991).
- Kobe, B. & Deisenhofer, J. Crystal-structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**, 751–756 (1993).
- Kobe, B. & Deisenhofer, J. A. Structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**, 183–186 (1995).
- Bentley, R. C. & Keene, J. D. Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B'' protein and through interactions with U2 snRNP-A' protein. *Mol. Cell. Biol.* **11**, 1829–1839 (1991).
- Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
- Tsai, D. E., Harper, D. S. & Keene, J. D. U1-snRNP-A protein selects a 10-nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res.* **19**, 4931–4936 (1991).
- William, D. J. & Hall, K. B. RNA hairpin with non-nucleotide spacers binds efficiently to the human U1A protein. *J. Mol. Biol.* **257**, 265–275 (1996).
- Boelens, W. C. *et al.* The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* **72**, 881–892 (1993).
- van Gelder, C. W. G. *et al.* A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *EMBO J.* **12**, 5191–5200 (1993).
- Jovine, L., Oubridge, C., Avis, J. M. & Nagai, K. Two structurally different RNA molecules are bound by the spliceosomal protein U1A by the same recognition strategy. *Structure* **4**, 621–631 (1996).
- Allain, F. H.-T., Howe, P. W. A., Neuhaus, D. & Varani, G. Structural basis of the RNA binding specificity of human U1A protein. *EMBO J.* **16**, 5764–5774 (1997).
- Jessen, T.-H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, 3447–3456 (1991).
- Harper, D. S., Fresco, L. D. & Keene, J. D. RNA-binding specificity of a *Drosophila* snRNP protein that shares sequence homology with mammalian U1-A and U2-B'' proteins. *Nucleic Acids Res.* **20**, 3645–3650 (1992).
- Polycarpou-Schwarz, M., Gunderson, S. I., KandelsLewis, S., Seraphin, B. & Mattaj, J. W. *Drosophila* SNF/D25 combines the functions of the 2 snRNP proteins U1A and U2B'' that are encoded separately in human, potato and yeast. *RNA* **2**, 11–23 (1996).
- Laird-Offringa, I. A. & Belasco, J. G. Analysis of RNA-binding proteins by *in vitro* genetic selection: identification of an amino-acid residue important for locking U1A onto its RNA target. *Proc. Natl Acad. Sci. USA* **92**, 11859–11863 (1995).
- Biswas, R., Wahl, M. C., Ban, C. & Sundaralingam, M. Crystal Structure of an alternating octamer r(GUAGUGUA)dc with adjacent G-U wobble pairs. *J. Mol. Biol.* **267**, 1149–1156 (1997).
- Avis, J. *et al.* Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structural stability and RNA binding. *J. Mol. Biol.* **257**, 398–411 (1996).
- Hall, K. B. Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry* **33**, 10076–10088 (1994).
- Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
- Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. The crystal structure of the estrogen receptor DNA binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* **75**, 567–578 (1993).
- Patikoglou, G. & Burley, S. K. Eukaryotic transcription factor–DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 289–325 (1997).
- Oubridge, C., Ito, N., Teo, H.-C., Fearnley, I. & Nagai, K. Crystallisation of RNA–protein complexes. II. The application of protein engineering for crystallisation of the U1A protein–RNA complex. *J. Mol. Biol.* **249**, 409–423 (1995).
- Price, S. R., Oubridge, C., Varani, G. & Nagai, K. in *RNA–Protein Interactions. A Practical Approach* (ed. Smith, C. W. J.) 37–74 (Oxford University Press, Oxford, 1998).
- Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **D50**, 760–763 (1994).
- de la Fortelle, E. & Brice, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Meth. Enzymol.* **276**, 472–494 (1997).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Crystallogr.* **D52**, 30–42 (1996).
- Jones, T. A. & Kjeldgaard, M. O. *The Manual* (Uppsala University, Sweden, 1994).
- Nicholls, K. A., Sharp, B. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Protein: Struct. Funct. Genet.* **11**, 281–296 (1991).
- Ensnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 133–138 (1997).

Acknowledgements. We thank W. Scott, E. de la Fortelle, J. Li, H. Teo, C. Oubridge and C. Kambach, L. Jovine and A. McCoy for their help; M. Perutz, A. Klug, D. Rhodes and J. Schwabe for critically reading the manuscript; G. Lingley, A. Lenton and K. Nagai for the illustrations; and staff at the ELETTRA, Daresbury and EMBL Hamburg synchrotrons for help with data collection. S.R.P. was supported by an MRC studentship. This project was funded by the MRC and Human Frontier Science Program.

Correspondence and requests for materials should be addressed to K.N. (e-mail: kn@mrc-lmb.cam.ac.uk). The coordinates have been deposited in the Protein Data Bank, accession code 1A9N.

Reconciling the spectrum of Sagittarius A* with a two-temperature plasma model

Rohan Mahadevan

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

The radio source Sagittarius A* is thought to be powered by gas accreting onto a supermassive black hole at the centre of our Galaxy^{1,2}. Using the high infrared accretion rates³, however, standard accretion models⁴ are unable to explain the observed low luminosity and spectral energy distribution^{5–8}, which has led to the consideration of a new model: advection-dominated accretion flows^{9–12}. In an advection-dominated flow, most of the accretion energy is stored as thermal energy in the gas which is then lost as the gas falls into the black hole. This model requires the protons to have a much higher temperature than the electrons, and the gas therefore has a two-temperature structure^{10,13,14}. Although this model explains the low total luminosity^{15–18} and much of the spectral energy distribution (from millimetre wavelengths to hard X-rays), it has been difficult to reconcile with low-frequency radio observations. Here we show that a neglected emission process associated with the protons naturally explains the radio observations without any ‘fine tuning’ of the model parameters. This result simultaneously supports the two-temperature model of the gas and suggests that an advection-dominated accretion flow onto a black hole of 2.5×10^6 solar masses provides an accurate description of Sagittarius A*.

Figure 1 shows the most up-to-date observations of the Galactic Centre¹⁸. The spectrum rises at radio and submillimetre frequencies $\nu \approx 10^7$ – 10^{12} Hz, where most of the emission occurs, and has a sharp drop in the infrared. The X-ray observations consist of a possible detection at soft X-ray energies, and firm upper limits in the hard X-rays. The X-ray error-box corresponds to uncertainties in the observed photon index which lies between 1.0 and 2.0 (ref. 18). At very high energies, the EGRET satellite has observed γ -ray emission from the Galactic Centre region⁸. But owing to the low angular resolution of the measurements, $\sim 1^\circ$, the observations should perhaps be considered as upper limits.

The spectrum from a two-temperature advection-dominated accretion flow (ADAF) is determined by the cooling properties of the protons and electrons in the flow. The protons are at virial temperatures at all radii (proton temperature $T_p \approx 10^{12}$ K close to the black hole) and cool by creating neutral pions¹⁹, while the electrons have much lower temperatures ($T_e \approx 10^{9.5}$ K) and cool by various optically thin processes, such as synchrotron, inverse Compton and bremsstrahlung radiation^{11,20}.

Figure 1a shows the spectrum from the ADAF model of Sgr A* in ref. 18. This spectrum fits the submillimetre to hard X-ray spectrum quite well, but fails to explain the non-uniform radio spectrum. The radio luminosity, L_ν , is well represented by $L_\nu \propto \nu^{0.2}$ up to $\nu \approx 43$ GHz, which subsequently rises to $L_\nu \propto \nu^{0.8}$ for $\nu \geq 86$ GHz (ref. 21). ADAF models of Sgr A* have always been unable to account for this break, and are substantially underluminous at frequencies below ~ 86 GHz. This poses a serious problem.

The observed excess of radio emission (beyond what the model predicts) has usually been attributed to a weak jet of material that might emerge from the ADAF; jets are known to be strong radio sources. High-resolution radio observations, however, have ruled this out^{22–24}, which severely constrains any outflow models. In this

case, a rather *ad hoc* electron-temperature profile might be needed to account for the excess radio emission¹⁸, which probably does not correspond to physical conditions. More importantly, recent high-resolution measurements constrain the actual size of the emitting region^{5,22}. These observations require large brightness temperatures (in excess of 10^{10} K) to explain the observed flux at 43 GHz and 86 GHz. In an ADAF, however, the electron temperature is always well below 10^{10} K at all radii¹¹, and therefore cannot account for these high temperatures.

This apparent problem is solved by considering another emission process associated with the protons. In addition to producing neutral pions, energetic proton collisions can also create charged pions, which subsequently decay into positrons and electrons (referred to here as $e^\pm$). This had been neglected in earlier work because these particles do not produce significant amounts of γ -ray

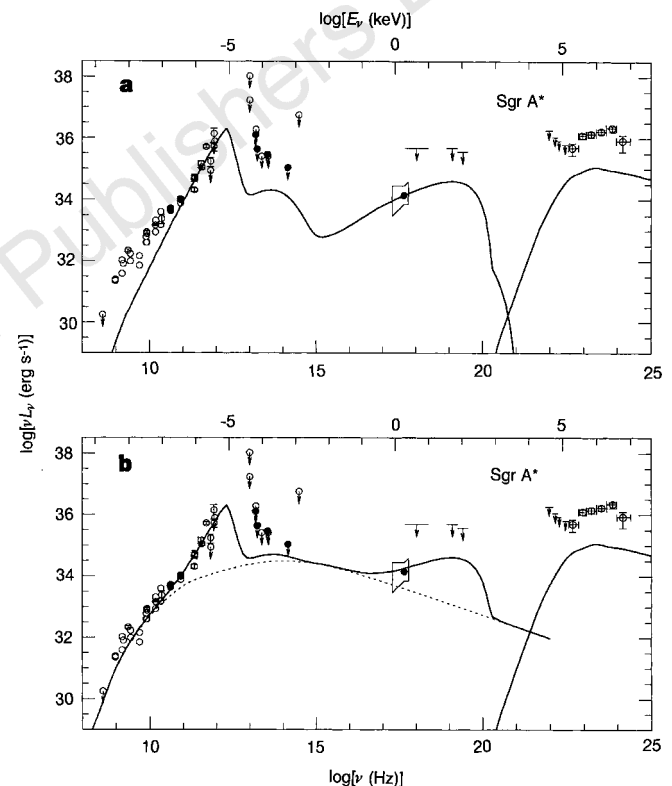


Figure 1 A comparison of the predicted emission from an ADAF model of the Galactic Centre with the observations. **a**, The spectrum of Sgr A*; the horizontal axis is the logarithm of the frequency and the vertical axis is the logarithm of the energy at that frequency. The measured fluxes were converted to luminosities assuming a distance of 8.5 kpc to the Galactic Centre. The data are the most up-to-date compilation of observations taken from ref. 18. The arrows represent upper limits, and the ‘box’ at frequency $\sim 10^{17}$ Hz represents the uncertainty in the observed photon index. The solid line is the spectrum from the baseline ADAF model of Sgr A* used in ref. 18. The ADAF parameters are $\alpha = 0.3$, $\beta = 0.5$, $M = 2.5 \times 10^6 M_\odot$, and $\dot{M} = 7.2 \times 10^{-6} M_\odot \text{ yr}^{-1}$, where α is the viscosity parameter⁴¹, β determines the strength of the magnetic field, and is defined so that $(1 - \beta)$ is the ratio of magnetic to total pressure, M is the dynamically measured mass of Sgr A*^{42,43}, and $\dot{M}$ is the mass accretion rate. For frequencies $\leq 10^{20}$ Hz, the spectrum is determined by the individual optically thin cooling processes of $\sim 10^{9.5}$ K thermal electrons, while for $\nu \geq 10^{20}$ Hz the spectrum is solely due to the decay of neutral pions. The discrepancy between the model and the observations above $\nu \sim 10^{20}$ Hz is not considered serious, as it is unclear at present whether the $\sim 1^\circ$ beam of EGRET is detecting a point source or some diffuse emission. These observations should therefore be considered as upper limits rather than detections of a central source. **b**, The solid line represents the total spectrum from the ADAF around Sgr A*, which includes the present results. The parameters used are identical to those in **a**. The dotted line represents only the synchrotron emission from the positrons and electrons.

emission¹⁹.

The high-energy $e^\pm$, however, can interact with the magnetic fields in the ADAF to produce synchrotron emission from radio to hard X-ray energies. Because the pions, and therefore the $e^\pm$, are created by proton–proton collisions, the energy spectra of the protons and $e^\pm$ are related. This allows a direct investigation of the assumption that the protons have a different average temperature from the electrons, and at the same time determines if the $e^\pm$ are created in sufficient number, and with the right energy, to produce the observed radio emission.

Here we assume that the energy spectrum of the protons is represented by a power-law distribution, $N(E_p) \propto E_p^{-s}$ with index s , where $N(E_p)$ represents the number of protons with energy E_p . The index is generally between 2 and 4, and we set it to $s = 2.75$, at the cosmic-ray value, suggesting that a similar acceleration mechanism might be at work in ADAFs¹⁹. The results are insensitive to the exact value of s (ref. 19).

The rate of production and energy spectrum of the $e^\pm$, $R(E)$ is determined by the frequency of proton collisions as well as their energy spectrum. For the assumed power-law proton distribution, the energy distribution of the $e^\pm$ is shown in Fig. 2. The spectrum rises at low energies, turns over at $E \approx 35$ MeV, and, as expected, extends as a power-law, E^{-s} , with the same energy dependence as the parent proton distribution²⁵. Because the created charged pion has a mass of ~ 140 MeV and decays into four particles, one of which is an electron or positron, we expect that on average the $e^\pm$ should carry away one-quarter of the total energy available (that is, $\sim 140/4 = 35$ MeV)²⁶. This is an expected turnover which is characteristic of $e^\pm$ production, and is shown in Fig. 2.

Determining the synchrotron emissivity from the $e^\pm$ requires a knowledge of their steady-state energy distribution $N(E)$. At a given energy E , the colliding protons produce $R(E)$ electrons and positrons. However, because the $e^\pm$ cool by synchrotron radiation, they lose their energy very efficiently, and the steady-state distribution is therefore determined by the competing effects of the creation

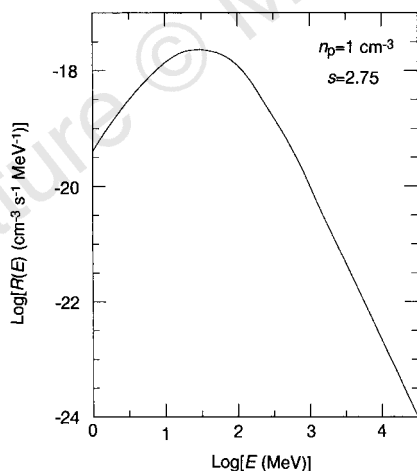


Figure 2 The energy spectrum, $R(E)$, of positrons and electrons that are created by colliding power-law protons with energy index $s = 2.75$. The vertical axis is the logarithm of number of positron and electrons created per unit volume, per second, per energy interval, and the horizontal axis is the logarithm of the energy. The scale on the vertical axis corresponds to a number density of protons equal to unity. For a number density N , the vertical axis must be multiplied by N^2 . The particles that are responsible for most of the emission are determined by the energy at which the function $E^2 R(E)$ peaks, which occurs in the range $100 \text{ MeV} < E < 500 \text{ MeV}$. The shape of the spectrum depends only on the physics of particle collisions and decays^{25,26}, and at high energies has the spectral shape $R(E) \propto E^{-s}$ (ref. 25). The spectrum therefore contains information about the parent proton distribution, as well as determining the shape of the resulting synchrotron spectrum. It therefore acts as a link between the form of the proton energy distribution and the observed synchrotron spectrum.

and depletion of particles. This requires the net flux of particles between two energies to be equal to their rate of injection, $d[N(E)\dot{E}_s(E)]/dE = R(E)$, where $\dot{E}_s(E)$ is the total synchrotron cooling rate as a function of energy²⁷.

Using the steady-state distribution $N(E)$, the $e^\pm$ synchrotron spectrum, from the ADAF around Sgr A*, is shown by the dotted line in Fig. 1b. The spectrum rises at low frequencies, turns over, and extends as a power-law at high frequencies. The spectral break at $\nu \approx 10^{15} \text{ Hz}$ is a direct consequence of the turn over in the $e^\pm$ energy spectrum shown in Fig. 2. At high frequencies, the spectrum is optically thin and has a spectral dependence, $L_\nu \propto \nu^{-s/2}$. The spectral slope therefore depends on the proton index s , which is a direct consequence of the $e^\pm$ having a steady-state distribution $N(E) \propto E^{-(s+1)}$ (ref. 27). At lower frequencies, the expected optically thin spectral dependence is $L_\nu \propto \nu^{-0.5}$ which corresponds to $N(E) \propto E^{-2}$ (ref. 27). However, in an ADAF, the emission at these low frequencies is self-absorbed by the plasma, and the resultant spectrum shown therefore has a different spectral dependence.

The solid curve in Fig. 1b represents the total radiation from the ADAF which includes this spectrum. At high frequencies $\approx 10^{13} \text{ Hz}$, the synchrotron emission contributes to, but does not significantly change, the total luminosity. In particular, the agreement with the X-ray flux is not affected, and the additional infrared flux is still well below the stringent upper limits.

At lower energies the result is striking. The emission reproduces the required spectral break at $\sim 86 \text{ GHz}$, is able to account for the 'excess' radio emission below this frequency, and diminishes sufficiently quickly at lower frequencies to agree with the radio upper limit at 400 MHz . As the emission at each radio frequency in Fig. 1a corresponds to a black-body spectrum at a given radius^{11,20}, the total spectrum shown by the solid line in Fig. 1b indicates that ADAFs produce more emission at a given frequency than the local black-body spectrum. The excess emission is from the high-energy electrons radiating at larger radii. This resolves the problem with the low energy radio emission: no outflow model is needed to account for the observed emission, and the high brightness temperatures inferred^{5,22} are easily accounted for by the non-thermal origin of the emission.

The quite good agreement with the radio observations suggests that the emission observed is most probably from the hot protons in the ADAF. But before drawing any conclusions, we examine the essential ingredients required to explain the radio spectrum. Assuming that the dynamics of the flow are determined, reproducing the radio spectrum requires high-energy electrons (or $e^\pm$) with energies $\sim 100 \text{ MeV}$ at all radii. In an ADAF, this requirement is naturally satisfied. Assuming that viscosity primarily heats the protons into a power-law distribution at all radii, the production of high-energy $e^\pm$ with the same energy is completely determined by only the nuclear physics of particle collisions and decays^{25,26}. In particular, the shape of the $e^\pm$ spectrum (compare Fig. 2) is fixed throughout the flow. We note that the number of $e^\pm$ produced is also the right amount; this is a natural consequence of the proton collision time being longer than the accretion time. Whereas shorter collision times would produce excessive amounts of $e^\pm$ which would result in too much radio emission, much longer collision times would result in too little radio emission.

The agreement of the theory with the observations depends on two basic assumptions of ADAFs that have always been debated: (1) the existence of a two-temperature plasma, and (2) that viscosity preferentially heats the protons. We have quite good observational evidence that the first assumption is probably true. This is because the radio to hard X-ray spectrum is determined by emission processes associated with both the protons and electrons, at their respective temperatures. If the temperatures of protons and electrons were the same, or were markedly different from their calculated values, the resulting spectrum would be completely different and fail to explain any of the observations.

The second assumption is supported by the present results, and can be discussed in terms of δ , which is the fraction of viscous energy that heats the electrons. The baseline model in ref. 18 set $\delta \approx 0.001$, and showed that for $\delta > 0.01$, too much radiation is produced, and the electron spectrum does not agree with the observations. Here we have a radiation mechanism that accounts for the other fraction $(1 - \delta)$ that heats the protons, and have shown that the agreement with the low-energy radio spectrum requires the amount of energy transferred to the electrons to be small. This shows that the average energy of the protons is probably virial.

Although past work has attempted to answer both these questions theoretically^{28–33}, the results here provide indirect observational evidence that these assumptions are probably valid. Further, theoretical models which reach contrary conclusions are probably based on assumptions that are not valid in ADAF^{18,34}. The present results could therefore be used as tools to aid future theoretical work in resolving these complex questions in plasma physics.

The present results have assumed that all the viscous energy is deposited into a power-law proton distribution, which might seem improbable. However, if half the viscous energy were transferred into a power-law distribution, and half into a thermal distribution, the number of $e^{\pm}$ created reduces only by a factor ~ 2 (ref. 19), and the results presented here do not change significantly. Therefore, although the agreement with the radio flux requires a power-law proton distribution, it does not require all of the viscous energy to be deposited into the power-law protons.

It is interesting that the good agreement with observations comes from a model in which both the viscous hydrodynamics and the radiative processes have been included self-consistently. Previous models that have attempted to explain the observed spectrum have been phenomenological^{35–37}, or made simplifying assumptions, such as ignoring the angular momentum of the accreting gas^{3,38,39}, or, as noted previously^{18,40}, have errors in the synchrotron calculation which renders the resulting spectrum suspect^{3,39}. The ADAF models therefore provides us with a unique self-consistent framework which enables accurate prediction of spectra from accreting black holes.

We stress that there is no fine tuning in the present results. Whereas previous work on ADAFs has not included the $e^{\pm}$ synchrotron radiation, the results presented here show that this process is essential to explaining the observed non-uniform radio spectrum. The model used is identical to that presented in ref. 18, and we have simply taken into account an additional physical process and emission mechanism in the two-temperature ADAF. It is remarkably that, using the same parameters as in ref. 18, an emission mechanism associated with the protons is able to naturally reproduce the entire radio spectrum including the observed spectral break at ~ 86 GHz. The agreement of the theory with the observations encourages us to take the natural explanation, and conclude that Sgr A* is in fact a 2×10^6 solar-mass black hole that is accreting by way of a two-temperature ADAF. $\square$

Received 23 March; accepted 19 May 1998.

- Mezger, P. G., Duschl, W. J. & Zylka, R. The Galactic Center: a laboratory for AGN? *Annu. Rev. Astron. Astrophys.* **7**, 289–388 (1996).
- Genzel, R., Hollenbach, D. & Townes, C. H. The nucleus of our Galaxy. *Rep. Prog. Phys.* **57**, 417–479 (1994).
- Melia, F. An accreting black hole model for Sagittarius A*. *Astrophys. J.* **387**, L25–L28 (1992).
- Frank, J., King, A. & Raine, D. *Accretion Power in Astrophysics* (Cambridge Univ. Press, 1992).
- Rogers, A. E. E. et al. Small-scale structure and position of Sagittarius A* from VLBI at 3 millimeter wavelength. *Astrophys. J.* **434**, L59–L62 (1994).
- Menten, K. M., Reid, M. J., Eckart, A. & Genzel, R. The position of Sagittarius A*: accurate alignment of the radio and infrared reference frames at the Galactic Center. *Astrophys. J.* **475**, L111–L114 (1997).
- Predehl, P. & Trümper, J. ROSAT observation of the Sgr A region. *Astron. Astrophys.* **290**, L29–L32 (1994).
- Merck, M. et al. Study of the spectral characteristics of unidentified galactic EGRET sources. Are they pulsar-like? *Astron. Astrophys. Suppl.* **120**, 465–469 (1996).
- Ichimaru, S. Bimodal behavior of accretion disks—Theory and application to Cygnus X-1 transitions. *Astrophys. J.* **214**, 840–855 (1977).
- Rees, M. J., Begelman, M. C., Blandford, R. D. & Phinney, E. S. Ion supported tori and the origin of radio jets. *Nature* **295**, 17–21 (1982).
- Narayan, R. & Yi, I. Advection-dominated accretion: underfed black holes and neutron stars. *Astrophys. J.* **452**, 710–735 (1995).
- Abramowicz, M. A., Chen, X., Kato, S., Lasota, J.-P. & Regev, O. Thermal equilibria of accretion disks. *Astrophys. J.* **438**, L37–L39 (1995).

- Shapiro, S. L., Lightman, A. P. & Eardley, D. M. A two-temperature accretion disk model for Cygnus X-1—Structure and spectrum. *Astrophys. J.* **204**, 187–199 (1976).
- Phinney, E. S. in *Plasma Astrophysics* (eds Guyenne, T. D. & Levy, G.) 337–341 (SP-161, ESA, Paris, 1981).
- Rees, M. J. in *The Galactic Center* (eds Riegler, G. R. & Blandford, R. D.) 166–176 (Am. Inst. Phys., New York, 1982).
- Narayan, R., Yi, I. & Mahadevan, R. Explaining the spectrum of Sagittarius A* with a model of an accreting black hole. *Nature* **374**, 623–625 (1995).
- Mammoto, T., Mineshige, S. & Kusunose, M. Spectrum of optically thin advection-dominated accretion flow around a black hole: application to Sagittarius A*. *Astrophys. J.* **489**, 791–803 (1997).
- Narayan, R., Mahadevan, R., Grindlay, J. E., Popham, R. G. & Gammie, C. Advection-dominated accretion model of Sagittarius A*: evidence for a black hole at the Galactic Center. *Astrophys. J.* **492**, 554–568 (1998).
- Mahadevan, R., Narayan, R. & Krolik, J. Gamma-ray emission from advection-dominated accretion flows around black holes: application to the Galactic Center. *Astrophys. J.* **486**, 268–275 (1997).
- Mahadevan, R. Scaling laws for advection-dominated flows: applications to low-luminosity galactic nuclei. *Astrophys. J.* **477**, 585–601 (1997).
- Falcke, H. et al. The simultaneous spectrum of Sgr A* from $\lambda 20$ cm to $\lambda 1$ mm and the nature of the mm-excess. *Astrophys. J.* (in the press).
- Backer, D. C. et al. Upper limit of 3.3 astronomical units to the diameter of the Galactic Center radio source Sgr A*. *Science* **262**, 1414–1416 (1993).
- Marcaide, J. M. et al. Position and morphology of the compact non-thermal radio source at the galactic center. *Astron. Astrophys.* **258**, 295–301 (1992).
- Alberdi, A. et al. VLA image of Sgr A* at $\lambda = 1.35$ cm. *Astron. Astrophys.* **277**, L1–L4 (1993).
- Ginzburg, V. L. & Syrovatskii, S. I. *The Origin of Cosmic Rays* (Macmillan, New York, 1964).
- Dermer, C. D. Binary collision rates of relativistic thermal plasmas. II. Spectra. *Astrophys. J.* **307**, 47–59 (1986).
- Rybicki, G. & Lightman, A. *Radiative Processes in Astrophysics* (Wiley, New York, 1979).
- Begelman, M. C. & Chiueh, T. Thermal coupling of ions and electrons by collective effects in two-temperature accretion flows. *Astrophys. J.* **332**, 872–890 (1988).
- Quataert, E. Particle heating by Alfvénic turbulence in hot accretion flows. *Astrophys. J.* (in the press); also as preprint astro-ph/9710127 (1998).
- Gruzinov, A. Radiative efficiency of collisionless accretion. *Astrophys. J.* (in the press); also as preprint astro-ph/9710132 (1998).
- Blackman, E. Fermi energization in magnetized astrophysical flows. *Phys. Rev. Lett.* (in the press); also as preprint astro-ph/9710137 (1998).
- Bisnovatyi-Kogan, G. S. & Lovelace, R. V. E. Influence of ohmic heating on advection-dominated accretion flows. *Astrophys. J.* **486**, L43–L46 (1997).
- Quataert, E. & Gruzinov, A. Turbulence and particle heating in advection-dominated accretion flows. *Astrophys. J.* (submitted); also as preprint astro-ph/9803112 (1998).
- Narayan, R., Mahadevan, R. & Quataert, E. Advection dominated accretion around black holes in *The Theory of Black Hole Accretion Discs* (eds Abramowicz, M. A., Björnsson, G. & Pringle, J. E.) (Cambridge Univ. Press, in the press).
- Duschl, W. J. & Lesch, H. The spectrum of Sgr A* and its variability. *Astron. Astrophys.* **286**, 431–436 (1994).
- Falcke, H. in *Unsolved Problems in the Milky Way* (eds Blitz, L. & Teuben, P. J.) 163–170 (IAU Symp. No. 169, Kluwer, Dordrecht, 1996).
- Becker, T. & Duschl, W. J. Synchrotron radiation from quasi-monoenergetic electrons. Modeling the spectrum of Sgr A*. *Astron. Astrophys.* **328**, 95–106 (1997).
- Mastichiadis, A. & Ozerov, L. M. X-ray and gamma-ray emission of Sagittarius A* as a wind-accreting black hole. *Astron. Astrophys.* **426**, 599–603 (1994).
- Melia, F. An accreting black hole model for Sagittarius A*. 2: A detailed study. *Astrophys. J.* **426**, 577–585 (1994).
- Mahadevan, R., Narayan, R. & Yi, I. Harmony in electrons: cyclotron and synchrotron emission by thermal electrons in a magnetic field. *Astrophys. J.* **465**, 327–337 (1996).
- Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
- Haller, J. et al. Stellar kinematics and the black hole in the Galactic Center. *Astrophys. J.* **456**, 194–205 (1996); Erratum. *Astrophys. J.* **468**, 955 (1996).
- Eckart, A. & Genzel, R. Stellar proper motions in the central 0.1 pc of the Galaxy. *Mon. Not. R. Astron. Soc.* **284**, 576–598 (1997).

Acknowledgements. I thank R. Narayan, E. Blackman, A. Fabian, C. Gammie, Z. Haiman, J. Herrnstein, J. Krolik, A. Mody, M. Rees and E. Quataert for discussions and comments.

Correspondence and requests for materials should be addressed to the author (e-mail: rohan@ast.cam.ac.uk).

Long-lived giant cells detected at the surface of the Sun

J. G. Beck*, T. L. Duvall Jr† & P. H. Scherrer*

* W. W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, California 94305-4085, USA

† Laboratory for Astronomy and Solar Physics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Giant convective cells have been predicted¹ to exist in the Sun. Such cells should span the entire zone unstable to convective motions—now known to cover the outer 29 per cent of the Sun's radius²—and could be dredging up the magnetic flux that is thought to be the source of solar activity (sunspots). Several studies^{3–5} have failed to detect these giant cells, although there

have been hints^{6–9} of their existence. We have detected long-lived velocity cells, which we identify as the elusive giant convective cells, extending over 40–50 degrees of longitude but less than 10 degrees of latitude. The large aspect ratio (>4) is surprising (although predicted by one model¹⁰) and may be a consequence of the Sun's differential rotation, whereby features with a larger extent in latitude are broken up by rotational shear.

Previous observations show that any giant-cell signal must show a velocity of less than 10 m s^{-1} (ref. 11). To see such a small signal amidst the cacophony of smaller-scale, higher-velocity signals requires careful observations and analysis. Continuous coverage and a very stable observing platform are crucial. To this end, we decided to study the giant-cell question using dopplergrams from the SOI/MDI instrument¹². The raw dopplergrams have $1,024 \times 1,024$ pixels covering the projected disk of the Sun, and are observed every minute; the primary objective of these measurements is helioseismic study. For the present work, we used dopplergrams to measure directly the flow of material at the surface. Unfortunately, the telemetry used to transmit the full-resolution dopplergrams to the ground is only available part of the time. However, low-resolution averages of the dopplergrams are computed each minute on board and later transmitted to Earth. For the present study, dopplergrams averaged into 140 bins covering most of the solar disk have been analysed for the period May 1996 to September 1997. We have been able to use data covering 97% of the time interval.

A more detailed description of the data reduction will be published elsewhere, so a brief discussion is presented here. The per-minute bin means are averaged over the 1,440 minutes in a day, leaving one set of 140 bins for each of 505 days to be analysed. To enable further analysis, some removal of extraneous signals is done: (1) the relative velocity of the spacecraft and each bin is subtracted,

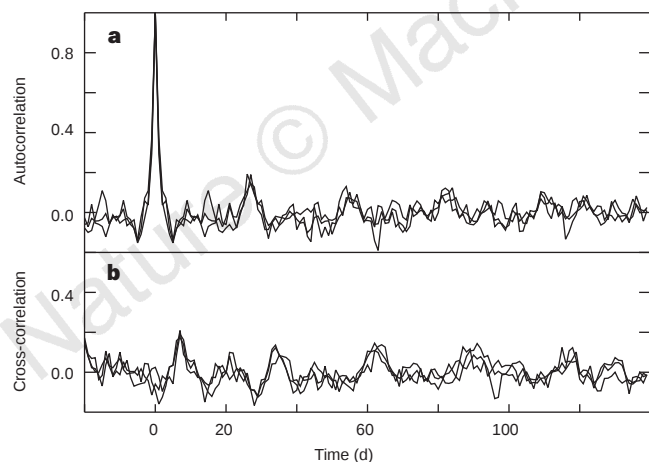


Figure 1 Autocorrelation (**a**) and cross-correlation (**b**) of the time series for selected bins. The size of the bins used for these figures is 10° in latitude by 15° in longitude. They are centred 5° north of the solar equator. The bins used for **a** are at 46° , 60° and 74° west longitude with respect to the central meridian. The slowly decaying features appear at multiples of the equatorial solar rotation period: 27 d, 54 d, 81 d and 108 d. These demonstrate the presence of long-lived structures rotating with the Sun. The correlation coefficient of ~ 0.2 near 27 d indicates that 20% of the variance is correlated after this period and 80% is not. The uncorrelated 80% is thought to arise from short-lived smaller convection cells. **b** shows the cross-correlation of time series for bins separated by 120° . Because a flow feature with a positive sense of Doppler shift on the east side will be seen as a negative Doppler shift 9 d later on the west side, the sign of the Doppler signals of bins at west longitudes is inverted. The bin pairs are: (46° east, 74° west), (60° east, 60° west), (74° east, 46° west). The peaks appearing in the correlation at 9 d, $9+27$ d, $9+54$ d, $9+81$ d and $9+108$ d are clear indications of long-lived horizontal flow features. The decay of the correlation coefficient is comparable to that seen in the autocorrelation plots.

(2) four discontinuities in the velocity signal caused by adjustments to the instrument set-up are corrected by addition of a constant after the adjustment, (3) some spurious spikes are removed, and (4) a large slow drift of the velocity signal is removed by fitting a seventh-order polynomial in time to the result of the previous steps. Most of the variance of the resultant signals is from horizontal motion, probably supergranulation. This can be seen from the root-mean-square (r.m.s.) signal varying from 2.5 m s^{-1} near the centre of the solar image to 10.2 m s^{-1} near the edge, where the system is more sensitive to horizontal motions.

To separate out the short-lived signals, we calculated the temporal autocorrelation function (Fig. 1a) for three bins. The peaks in this function at multiples of the solar rotation period of 27 d clearly show the existence of long-lived features. The decay in the height of this peak suggests a lifetime of four months for the causative signals. The height of the first peak at 27 d shows that 0.2 of the variance repeats after this interval, or an r.m.s. signal of 3.6 m s^{-1} . The full-width at half-maximum (of 3.3 days of the peak near 27 d suggests that the size of the features is $(3.3 \text{ d}) \times (13.3^\circ \text{ d}^{-1})$, or 44° in longitude, where $13.3^\circ \text{ d}^{-1}$ is the solar rotation frequency at this latitude. Clear signatures of long-lived features have been detected in this way in the latitude range $\pm 30^\circ$.

In a previous search¹³ for giant cells, spurious Doppler shifts from sunspot regions masqueraded as a long-lived signal. A characteristic of such a signal is that it would appear symmetric on the east and west sides of the solar disk. A true giant-cell signal would probably have a component that is antisymmetric across the centre of the disk. That is, a horizontal motion towards the west would show blue shift when seen on the east side of the image and later, when it rotated to the west side, it would show a red shift. To look for this component, we calculated the cross-correlations of the signal from the bins mentioned in the previous paragraph with counterparts on

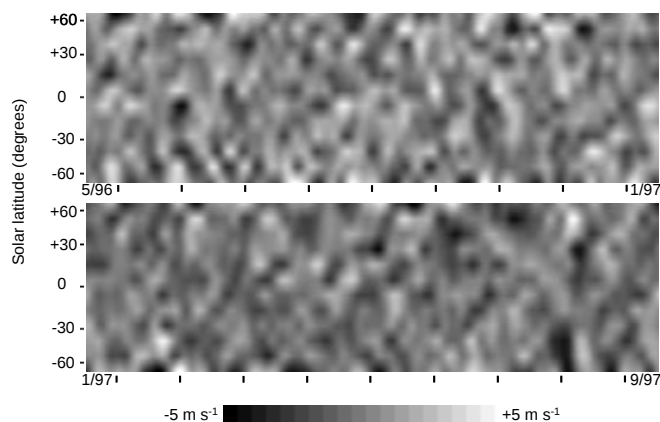


Figure 2 Maps of the east-west component of flow speed obtained by averaging over a disk passage. The signals are weighted by the sine of the longitude and are shifted in time before averaging to account for the solar rotation. A white signal corresponds to flow in the direction of solar rotation. Time is on the horizontal axis, with the first half of the 505 d covered in the top image and the second half in the image below. The range of dates covered is shown in the lower corners of the images. The small black boxes along the bottom of the images are separated by the rotation period defined by Carrington of 27.2753 d. About half the variance in these images is due to the long-lived structures; the other half of the variance is caused by shorter-lived smaller-scale supergranules.

the west side of the central meridian (Fig. 1b). It takes ~ 9 d for a feature to rotate from one bin to the other. So if there is a long-lived east–west horizontal motion, it will reveal itself in Fig. 1b by a correlation near 9 d, and then also at $9 + 27$, $9 + 54$, etc. This is what we see in Fig. 1b, confirming that we have detected long-lived, giant features. Such distinctive signatures have been seen in the latitude range $\pm 30^\circ$.

To reduce the effect of short-lived noise, we derived weighted averages of the data from all longitudes with the same central latitude to isolate the signal antisymmetric across the central meridian. The resultant map is shown in Fig. 2. Features repeating at the time interval of one solar rotation are evident. An autocorrelation of the same latitude band ($5 \pm 5^\circ$) as Fig. 1 is shown in Fig. 3. The amplitude of the long-lived component is clearly larger; this is the effect of averaging over several lifetimes of the shorter-lived supergranular signal. Using the data in Fig. 3, we calculate that the long-lived component makes up half the variance in the map of Fig. 2.

We have tried to measure the latitudinal extent of the long-lived features by examining the cross-correlations between bins of identical longitude and adjacent in latitude at a lag of 27 d. This correlation is very small compared to the peak in the autocorrelation of 0.2 at 27 d. From a simple linear interpolation, we then estimate a full-width at half-maximum of 10° in latitudinal extent, which is the latitude spacing of our data points. This, of course, is just an upper limit to the size. From our earlier measurement of longitudinal extent of 44° , we conclude the aspect ratio is >4 . This is perhaps the most surprising result of this analysis. Some models predict exactly the opposite¹⁴, that cells would be elongated in latitude and be ‘banana-shaped’. These models also predicted that the interior rotation rate would be constant on cylinders, whereas it is known now from helioseismology that the rotation is approximately constant on radii in the convective zone^{15,16}. It seems that long-lived features extended in latitude would be a natural consequence of rotation being constant on cylinders, while rotation constant on radii but varying with latitude might only allow long-lived features with a narrow latitudinal extent. In essence, the rotational shear may break up features that would otherwise be larger in extent. The modelling of convection in the Sun has become increasingly sophisticated in recent years^{17–19}. We might expect in future years to be able to model convection cells with the large aspect ratio measured here.

We have measured the scale of the features in our correlation functions using the full-width at half-maximum. It would be useful to relate this to the Fourier spectrum of the convection, which is used in some theories. In the Fourier representation, the convection is made up of components that have a harmonic form as $\cos(m\phi)$, where ϕ is the longitude and m is the number of waves in a

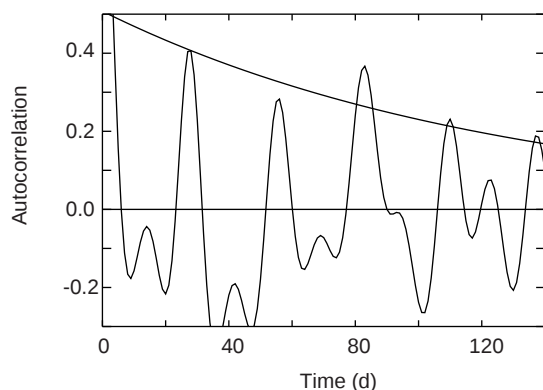


Figure 3 The autocorrelation computed for the map of Fig. 2 for the latitude $5 \pm 5^\circ$. The signals at multiples of the solar rotation period have higher amplitudes, as more of the short-lived noise has been averaged out. Also shown is an exponential fit to the peak amplitude. The fit falls to e^{-1} of its original value in 126 d.

circumference which only takes on integer values. For a spectrum with a single m , the correlation function is also harmonic and so our measurement of full-width at half-maximum can be related directly to m . We find that our width of 44° in longitude corresponds to $m \approx 3$. We imagine that the convection is made up of a number of components of different m and different amplitudes. With a spectrum made up of a number of components, our measurement corresponds to a mean value weighted by the variance. The present analysis is sensitive to cells with $m < 14$, although some of the higher values will be reduced in amplitude by the solar rotation. In future work, it would be useful to measure the Fourier spectrum to get the distribution function of the convection.

An important aspect of future work would be to uncover the relation of giant cells to solar activity. Solar activity seems to persist at the same location for much longer than is reasonable as a chance occurrence²⁰. It is thought that the magnetic fields of solar activity are generated at the bottom of the convection zone; so convection cells connected to this layer that last for months could provide a natural explanation for the persistence of solar activity. It may be that the giant cell is transporting to the surface the magnetic flux that is generated at the bottom of the convective layer. We also need to resolve the latitudinal size of the cells; we hope that this could be done using the high-resolution averages of the raw dopplergrams (0.6° bin size) that are transmitted to Earth each minute. Do these cells extend to the bottom of the convective zone, as originally predicted¹? This question could be addressed by use of the technique of time–distance helioseismology, which has recently detected subsurface flows^{21–23}. □

Received 11 March; accepted 19 May 1998.

- Simon, G. W. & Weiss, N. O. Supergranules and the hydrogen convection zone. *Z. Astrophys.* **69**, 435–450 (1968).
- Christensen-Dalsgaard, J., Gough, D. O. & Thompson, M. J. The depth of the solar convection zone. *Astrophys. J.* **378**, 413–437 (1991).
- LaBonte, B. J., Howard, R. & Gilman, P. A. An improved search for large-scale convection cells in the solar atmosphere. *Astrophys. J.* **250**, 796–798 (1981).
- Snodgrass, H. B. & Howard, R. Limits on photospheric Doppler signatures for solar giant cells. *Astrophys. J.* **284**, 848–855 (1984).
- Chiang, W.-H., Petro, L. D. & Foukal, P. V. A photometric search for solar giant convection cells. *Sol. Phys.* **110**, 129–138 (1987).
- Cram, L. E., Durney, B. R. & Guenther, D. B. Preliminary observations of velocity fields at the solar poles. *Astrophys. J.* **267**, 442–454 (1983).
- Hathaway, D. H. et al. GONG observations of solar surface flows. *Science* **272**, 1306–1309 (1996).
- Simon, G. W. & Strous, L. H. Supergranular evolution, solar rotation, and a search for giant cells. *Bull. Am. Astron. Soc.* **29**, 1402 (1997).
- Ulrich, R. K. Large scale convection and the solar radius. In *New Eyes to See Inside the Sun and Stars* (eds Deubner, F.-L., Christensen-Dalsgaard, J. & Kurtz, D. W.) (Proc. IAU Symp. 185, Kluwer, in the press).
- Snodgrass, H. B. & Wilson, P. R. Solar torsional oscillations as a signature of giant cells. *Nature* **328**, 696–699 (1987).
- Brown, T. M. & Gilman, P. A. Techniques for detecting giant cells using spatially resolved solar velocity data. *Astrophys. J.* **286**, 804–809 (1984).
- Scherrer, P. H. et al. The solar oscillations investigation—Michelson Doppler imager. *Sol. Phys.* **162**, 129–188 (1995).
- Scherrer, P. H., Bogart, R., Hoeksema, J. T. & Yoshimura, H. in *Seismology of the Sun and Distant Stars* (eds Gough, D. O.) 93–102 (Reidel, Dordrecht, 1986).
- Gilman, P. A. Nonlinear boussinesq convective model for large scale solar circulations. *Sol. Phys.* **27**, 3–26 (1972).
- Duvall, T. L. Jr, Harvey, J. W. & Pomerantz, M. A. Latitude and depth variation of solar rotation. *Nature* **321**, 500–501 (1986).
- Schou, J. et al. Helioseismic studies with SOI-MDI of differential rotation in the solar envelope. *Astrophys. J.* (in the press).
- Nordlund, Å. & Dravins, D. Stellar granulation. III. Hydrodynamic model atmospheres. *Astron. Astrophys.* **228**, 155–183 (1990).
- Kim, Y.-C., Fox, P. A., Sofia, S. & Demarque, P. Modeling of shallow and inefficient convection in the outer layers of the sun using realistic physics. *Astrophys. J.* **442**, 422–433 (1995).
- Freytag, B., Ludwig, H.-G. & Steffen, M. Hydrodynamical models of stellar convection; the role of overshoot in DA white dwarfs, A-type stars, and the sun. *Astron. Astrophys.* **313**, 497–516 (1996).
- Bai, T. Distribution of flares on the sun during 1955–1985; ‘hot spots’ (active zones) lasting for 30 years. *Astrophys. J.* **328**, 860–878 (1988).
- Duvall, T. L. Jr, D’Silva, S., Jefferies, S. M. & Harvey, J. W. Downflow under sunspots detected by helioseismic tomography. *Nature* **379**, 235–237 (1996).
- Kosovichev, A. G. Tomographic imaging of the sun’s interior. *Astrophys. J.* **461**, L55–L57 (1996).
- Giles, P. M. et al. A subsurface flow of material from the sun’s equator to its poles. *Nature* **390**, 52–54 (1997).

Acknowledgements. T.L.D. thanks P. Scherrer and the SOI group at Stanford for their hospitality during this work and for the use of their computing facilities, supported by NASA. SOHO is a mission of international cooperation between ESA and NASA. This work was supported in part by the Solar Physics Branch of the Space Physics Division of NASA.

Correspondence and requests for materials should be addressed to J.G.B. (e-mail: jbeck@sol.stanford.edu).

Visualization of hydrogen migration in solids using switchable mirrors

F. J. A. den Broeder*, S. J. van der Molen†, M. Kremers†, J. N. Huijberts†, D. G. Nagengast†, A. T. M. van Gogh†, W. H. Huisman†, N. J. Koeman†, B. Dam†, J. H. Rector†, S. Plota†, M. Haaksma†, R. M. N. Hazen*, R. M. Jungblut*, P. A. Duine* & R. Griessen†

* Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

† Institute COMPAS and Faculty of Physics and Astronomy, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

Switchable mirrors^{1–3} made of thin films of the hydrides of yttrium (YH_x), lanthanum (LaH_x) or rare-earth metals exhibit spectacular changes in their optical properties as *x* is varied from 0 to 3. For example, α-YH_{x<0.23} is a shiny, hexagonally close-packed metal, β-YH_{2±δ} is a face-centred cubic metal with a blue tint in reflection and a small transparency window at red wavelengths, whereas hexagonally close-packed γ-YH_{x>2.85} is a yellowish transparent semiconductor. Here we show that this concentration dependence of the optical properties, coupled with the high mobility of hydrogen in metals, offers the possibility of real-time visual observation of hydrogen migration in solids. We explore changes in the optical properties of yttrium films in which hydrogen diffuses laterally owing to a large concentration gradient. The optical transmission profiles along the length of the film vary in such a way as to show that the formation of the various hydride phases is diffusion-controlled. We can also induce electromigration of hydrogen, which diffuses towards the anode when a current flows through the film. Consequently, hydrogen in insulating YH_{3–δ} behaves as a negative ion, in agreement with recent strong-electron-correlation theories^{4,5}. This ability to manipulate the hydrogen distribution (and thus the optical properties) electrically might be useful for practical applications of these switchable mirrors.

Diffusion of hydrogen in metals has attracted considerable attention^{6–8}, because it is fast, involves only simple jumps between interstitial sites, and is ideal for the investigation of isotope effects when H is replaced by deuterium, tritium or the much lighter positive muon.

The central idea of our present work is to exploit the characteristic features of the transmission spectra of switchable mirrors to monitor hydrogen diffusion optically in a non-invasive way. Because the relative hydrogen concentration *x* can be as high as 3 and the phase diagrams of these metal hydrides consist of several phases⁹, this provides us with the possibility of investigating hydrogen diffusion in such systems. At moderate temperatures (for example, 400 K) diffusion is already fast enough to be observed in real-time with the naked eye or with standard optical equipment.

Our samples are produced in the following way. First, a 300-nm-thick yttrium film is evaporated under ultrahigh vacuum conditions onto a transparent substrate (sapphire or silica) by means of an electron gun. Subsequently, a 30-nm-thick palladium pattern (consisting of a disk or one or more strips) is evaporated *in situ* on top of the yttrium. In air, the yttrium oxidizes, forming a thin Y₂O₃ layer, which is impermeable to hydrogen atoms. However, areas covered with Pd do not oxidize, and remain permeable for hydrogen.

In a typical experiment, hydrogen gas (10⁵ Pa) is introduced into a chamber containing the sample. The chamber is equipped with optical windows and a temperature control system and can be placed onto the positioning table of an optical microscope (Olympus BX60F5). Using a white lamp behind the cell, optical transmis-

sion changes are monitored by means of a three-CCD red-green-blue (RGB) colour camera (Sony DXC-g50P). In contact with H₂, the yttrium underneath the palladium pattern immediately starts absorbing hydrogen atoms, because Pd is an excellent catalyst for H₂ dissociation. So within a few seconds, transparent YH_{3–δ} is formed in the Pd-covered region. Further hydrogen uptake can only take place if H diffuses out laterally, that is, into the Y underneath the transparent Y₂O₃ layer.

In Fig. 1b–d we show the time evolution of radial hydrogen diffusion using the circular geometry shown in Fig. 1a. As the hydrogen concentration *x* decreases from essentially 3 beneath the Pd to 0 far from it, several hydride phases are formed. As each phase has its characteristic transmission, H diffusion leads to the forma-

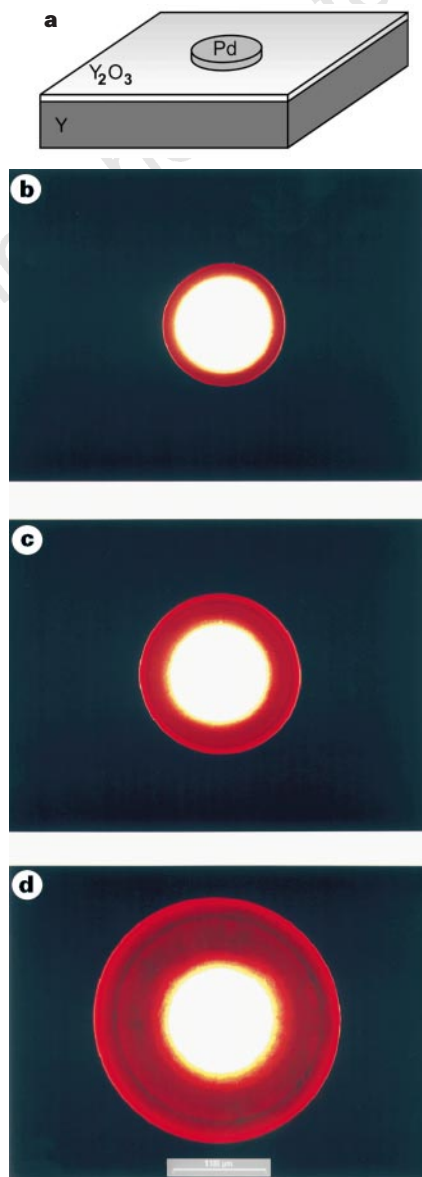


Figure 1 Radical hydrogen diffusion in yttrium. Microscope photographs in transmitted light of a 300-nm-thick yttrium film on sapphire, covered by a 30-nm-thick palladium disk of 1.1 mm diameter (a) at various times *t* after introduction of H₂ (10⁵ Pa). Within a few seconds, transparent γ-YH_{3–δ} is formed beneath the Pd, which adsorbs and dissociates hydrogen. The remaining Y surface has been oxidized, making it impermeable for hydrogen. Diffusion profiles are shown after 1 day (b), after 5 days (c), and after 23 days (d) at 60°C. The red outer ring is due to the weak transparency window at *hν* = 1.8 eV, characteristic of the dihydride β-YH_{2±δ} phase, while the dark surrounding region is non-transparent, metallic α-YH_x. Scale bar in d, 1100 μm.

tion of typical rings which expand with time. The inner ring in Fig. 1b–d corresponds to the γ -YH_{3-δ} phase, whereas the outer ring corresponds to the β -YH_{2±ε} phase, with its weak transparency window for red photons ($\hbar\omega = 1.8$ eV). We note that it is the optical features in YH_x (see Fig. 4 in ref. 10) that lead to such spectacular effects, making it possible to observe real-time hydrogen diffusion in a solid at room temperature. Moreover, this also provides the possibility of observing hydrogen diffusion in other materials, by using a switchable mirror layer evaporated on top of the sample as an optical indicator of local hydrogen presence.

For a scaling analysis of the diffusion process, it is necessary to consider the one-dimensional linear diffusion situation shown in Fig. 2a. Fick's second law is then:

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial z} \left(D(c) \frac{\partial c}{\partial z} \right) \quad (1)$$

where z denotes the distance measured from the Pd strip, t is time, and c is the hydrogen concentration; c is related to $x \equiv [H]/[Y]$, the yttrium molar volume V_M^Y , and Avogadro's number N_A by

$c = (xN_A)/(V_M^Y(x))$. The H diffusion coefficient $D(c)$ is explicitly allowed to be concentration dependent.

Boltzmann¹¹ (see also ref. 12) showed that equation (1) can be transformed into a relation in one variable $\lambda \equiv z/\sqrt{t}$ if the boundary conditions can be transformed accordingly. Moreover, this result still holds in the case of multiphased systems^{13,14} and yields:

$$\lambda \frac{dc}{d\lambda} = -2 \frac{d}{d\lambda} \left(D(c) \frac{dc}{d\lambda} \right) \quad (2)$$

The boundary condition $c(0, t) = c_0$ is easily satisfied by requiring $c(\lambda = 0) = c_0$, whereas $c(\infty, t) = 0$ translates into $c(\lambda = \infty) = 0$. Equation (2) implies that there are so-called similarity solutions: $c(z, t) = c(\lambda)$ or, equivalently, that all curves $c(z, t)$ should fall onto each other when transformed to $c(z/\sqrt{t})$. As the optical transmission T_{opt} is a function of c , we can check the validity of Boltzmann scaling by plotting $T_{\text{opt}}(z/\sqrt{t})$. Fig. 2b–d shows three RGB photographs for linear diffusion at 100 °C. The CCD R-signal output obtained from Fig. 2d is shown in Fig. 2e. The sharp discontinuities in Fig. 2e and f correspond respectively to the α - β , and β - γ phase boundaries of the YH_x system. The 'noise' in the signal in the γ -phase is due to the extreme sensitivity of the optical response to spatial concentration variations near the metal–insulator transition. It can thus be seen as a 'fingerprint' of minute fluctuations in the homogeneity of the film. In Fig. 3, the R and B profiles recorded at 16 different times are plotted as a function of λ . Except for the first curve, the profiles scale remarkably well, proving that the Boltzmann transformation applies to our system. This implies that the entire process is diffusion-controlled, and that all stable hydride phases are present in the sample¹⁵. Consequently, any spatially resolved concentration profile gives a 'bird's eye view' of the entire phase diagram of the YH_x-system. Not only is this valuable for the unravelling of the complicated phase diagrams of all rare-earth hydrides, but it also provides the possibility of measuring the concentration dependence of various physical properties (for example, resistivity) simply by doing local measurements.

Another important consequence of the validity of Boltzmann scaling is the possibility of determining the diffusion coefficient $D(c)$ as a function of concentration from a single profile. This follows directly from equation (2), which, after integration between $\lambda = \infty$ (at $z = \infty$) and a chosen λ_c gives the general result:

$$D(c) = - \frac{\int_0^{c(\lambda_c)} \lambda dc}{2 \left. \frac{dc}{d\lambda} \right|_{\lambda=\lambda_c}} \quad (3)$$

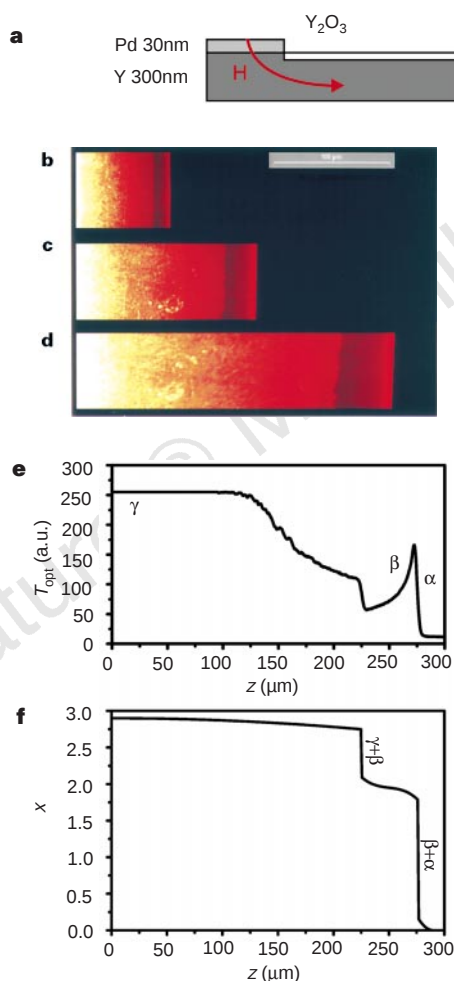


Figure 2 Linear hydrogen diffusion in yttrium. Linear one-dimensional hydrogen diffusion at 100 °C in a 300-nm-thick Y film on silica (a), covered on the left by a 30-nm Pd strip. b, Picture taken to the immediate right of the Pd strip at time $t = 1$ h after H₂ loading, c, at $t = 3$ h 24 min, and d, at $t = 10$ h 39 min. Panel e gives the intensity profile of the red (R) output signal of the three-CCD camera from d. At low z , the R-signal saturates. The various phases of the YH_x system are indicated by α , β and γ . The sharp optical discontinuities correspond to the α - β and β - γ phase boundaries or, equivalently, to vanishingly narrow mixed phase regions. This is indicated in panel f, which represents the hydrogen concentration profile $x(z)$ qualitatively. Scale bar, 100 μm .

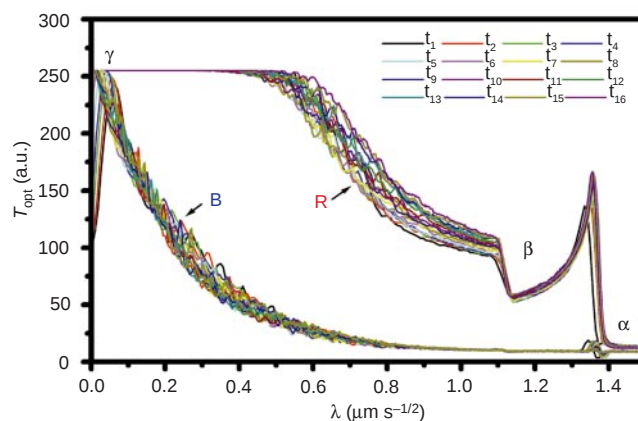


Figure 3 Scaling analysis. Boltzmann scaling of the optical transmission profiles for the sample in Fig. 2 at 16 different times (seconds) $t_i = 2,986, 3,592, 4,230, 4,940, 5,850, 7,000, 8,100, 10,120, 12,210, 14,610, 17,010, 20,310, 23,610, 27,030, 30,630$ and $38,310$ ($i = 1, \dots, 16$) by doing the transformation $\lambda = z/\sqrt{t_i}$. Both the red (R) and blue (B) signals of the three-CCD camera are plotted to demonstrate scaling for all λ .

The λ -dependence of the concentration, $c(\lambda)$, can be obtained from the optical transmission $T_{\text{opt}}(\lambda)$ in the diffusion samples, together with $T_{\text{opt}}(c)$ measured separately in homogeneously loaded samples. Because a complete mapping of $T_{\text{opt}}(c)$ is not yet available, we will restrict ourselves to the effective mobilities of the phase boundaries (defined by $K^{\text{ij}} = z_{\text{ij}}^2/t$, where $z_{\alpha\beta}$ ($z_{\beta\gamma}$) denotes the position of the $\alpha\beta$ ($\beta\gamma$) front), which determine the switching time of devices. Measurements of temperature dependence in the range $290 \text{ K} < T < 430 \text{ K}$ reveal perfect Arrhenius behaviour¹³; that is $K(T) = K_0 \exp(-E_a/kT)$. We find:

$$K_0^{\alpha\beta} = (3.1 \pm 0.6) \times 10^{-3} \text{ cm}^2 \text{ s}^{-1}, \quad E_a^{\alpha\beta} = 0.383 \pm 0.006 \text{ eV}$$

$$K_0^{\beta\gamma} = (1.4 \pm 0.4) \times 10^{-3} \text{ cm}^2 \text{ s}^{-1}, \quad E_a^{\beta\gamma} = 0.369 \pm 0.009 \text{ eV}$$

These results are consistent with literature values⁹ for the β -phase ($0.34 < E_a < 0.43 \text{ eV}$). For the γ -phase, only a few values (0.86 eV for $\text{YH}_{2.83}$ and 0.24 eV for $\text{YH}_{2.93}$) have been reported¹⁶.

Our optical method can also be applied to the case of hydrogen migration in the presence of an electric field. In such electromigration¹⁷ experiments, the motion of hydrogen is influenced in two ways. First, the field can exert a direct force on H; second, the charge carriers transfer momentum to hydrogen atoms, which act as scattering centres, resulting in the so-called wind force. In $\gamma\text{-YH}_{3-\delta}$, one expects the direct force to dominate because the electrical resistivity is relatively high¹⁷. Electromigration measurements can thus provide information about the electronic state of hydrogen in the γ -phase.

The samples used for this study (see Fig. 4a) are similar to that in

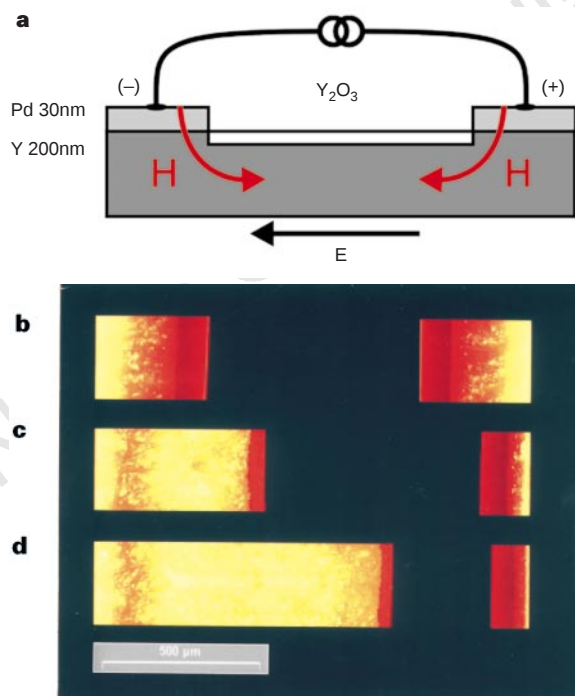


Figure 4 Electromigration of hydrogen in yttrium. Electromigration of hydrogen in 200-nm-thick, 1.5-mm-wide Y films on sapphire, covered on the left and right by 30-nm-thick Pd-strips (a). At $t = 0$, H_2 gas (10^5 Pa) is introduced, and hydrogen is allowed to migrate for 12 h 10 min ($t = t_1$) at 110°C . In the absence of an electric current (b), the resulting diffusion profiles at t_1 are symmetric. c and d show the hydrogen fronts at t_1 when a d.c. current ($I = 20 \text{ mA}$ and 40 mA , respectively) has been applied from right (+) to left (−). The asymmetry in the profiles is due to the force created by the electric field on the hydrogen atoms, and proves the negative effective valence of H in $\text{YH}_{3-\delta}$. We note that the effect of electromigration on the β -phase growth is relatively small due to its lower resistivity. Therefore, it grows faster than $\text{YH}_{3-\delta}$ at the positive side of the sample, but shrinks to minimal size at the negative side, where it is almost overtaken by the accelerated $\text{YH}_{3-\delta}$ phase. Scale bar, $500 \mu\text{m}$.

Fig. 2, but are covered by palladium at both ends. In Fig. 4b–d, we show photographs of hydrogen-migration patterns for samples that have been in a 1 bar H_2 atmosphere at 110°C for 12 hours and 10 minutes in three different situations. Figure 4b shows the diffusion pattern in the absence of an electric current. As expected, the fronts are symmetric with respect to the centre of the sample. Figure 4c,d shows H migration in the presence of a 20-mA and 40-mA current, respectively. The migration of the γ -phase on the side of the negative electrode (left-hand side) is much faster than on the positive (right) side. This implies that hydrogen in $\text{YH}_{3-\delta}$ behaves as a negatively charged impurity. This important result is in agreement with the theoretical views of Ng *et al.*⁴ and Eder *et al.*⁵, who propose that strong electron correlation effects determine the electronic properties of switchable mirrors. In particular, it is consistent with the formation of a sort of Zhang–Rice singlet^{4,18} binding two electrons¹⁹ (instead of two holes in the analogous case of the superconducting copper oxides) around each proton. The negative valence of hydrogen in $\text{YH}_{3-\delta}$ is in sharp contrast with the positive values found in metal–hydrogen systems such as VH_x , NbH_x , TaH_x and PdH_x , where hydrogen behaves essentially as a bare proton^{20,21}.

The large electromigration effect is also technologically important. As the γ -phase shrinks again on current reversal (not shown), reversible electric manipulation of the hydrogen distribution in these films is possible. For practical applications, it is of course the control of the hydrogen distribution perpendicular to the substrate that will be relevant. As the contrast between the optically closed ($\text{REH}_{2\pm e}$; RE, rare-earth) and open state ($\text{REH}_{3-\delta}$) can be as large as 10^7 (P.A.D. *et al.*, manuscript in preparation), relatively thin films could be used to achieve switching times far below 1 second. The fact that alloys such as $\text{Y}_{1-y}\text{La}_y\text{H}_x$ (A.T.M.v.G. *et al.*, manuscript in preparation), $\text{Y}_{1-y}\text{Mg}_y\text{H}_x$ (D.G.N. *et al.*, manuscript in preparation) and $\text{Gd}_{1-y}\text{Mg}_y\text{H}_x$ (ref. 22) also exhibit (in the latter cases, colour-neutral) switchable properties offers a wide range of possibilities for further fine tuning.

Received 26 March; accepted 1 June 1998.

- Huiberts, J. N. *et al.* Yttrium and lanthanum hydride films with switchable optical properties. *Nature* **380**, 231 (1996).
- Huiberts, J. N. *et al.* Synthesis of yttrium trihydride films for ex-situ measurements. *J. Alloys Compounds* **239**, 158 (1996).
- Griessen, R. *et al.* Yttrium and lanthanum hydride films with switchable optical properties. *J. Alloys Compounds* **253–254**, 44 (1997).
- Ng, K. K., Zhang, F. C., Anisimov, V. I. & Rice, T. M. Electronic structure of lanthanum hydrides with switchable optical properties. *Phys. Rev. Lett.* **78**, 1311 (1997).
- Eder, R., Pen, H. A. & Sawatzky, G. A. Kondo-lattice-like effects of hydrogen in transition metals. *Phys. Rev. B* **56**, 10115 (1997).
- Völkl, J. & Alefeld, G. Diffusion of hydrogen in metals. *Topics Appl. Phys.* **28**, 321 (1978).
- Fukai, Y. & Sugimoto, H. Diffusion of hydrogen in metals. *Adv. Phys.* **34**, 263 (1985).
- Qi, Zh., Völkl, J., Lässer, R. & Wenzl, H. Tritium diffusion in V, Nb and Ta. *J. Phys. F* **13**, 2053 (1983).
- Vajda, P. in *Handbook on the Physics and Chemistry of Rare Earths* Ch. 20 (eds Gschneider, K. A. & Eyling, L.) 207–291 (Elsevier, Amsterdam, 1995).
- Kremers, M. *et al.* Optical transmission spectroscopy of switchable yttrium hydride films. *Phys. Rev. B* **57**, 4943 (1998).
- Boltzmann, L. Zur Integration der Diffusionsgleichung bei variablen Diffusionskoeffizienten. *Weid. Ann.* **53**, 959 (1894).
- Crank, J. *The Mathematics of Diffusion* 105–106 & 230–232 (Clarendon, Oxford, 1975).
- Kidson, G. V. Some aspects of the growth of diffusion layers in binary systems. *J. Nucl. Mater.* **3**, 21 (1961).
- Adda, Y. & Philibert, J. *La Diffusion dans les Solides* 614–621 (Inst. Nat. des Sciences et Techniques Nucléaires, Saclay, 1966).
- Tu, K.-N., Mayer, J. W. & Feldman, L. C. *Electronic Thin Film Science* 294–324 (Macmillan, New York, 1992).
- Weaver, H. T. Nuclear magnetic resonance study of hydrogen diffusion in yttrium trihydride. *J. Chem. Phys.* **56**, 3193 (1972).
- van Ek, J. & Lodder, A. Electromigration of hydrogen in metals: theory and experiment. In *Defect and Diffusion Forum* Vols 115, 116, 1 (1994).
- Zhang, F. C. & Rice, T. M. Effective hamiltonian for the superconducting Cu oxides. *Phys. Rev. B* **37**, 3759 (1988).
- Griessen, R. Schaltbare Spiegel aus Metallhydriden. *Phys. Bl.* **53**, 1207 (1997).
- Verbruggen, A. H., Griessen, R., Rector, J. H. Hall voltage induced by hydrogen diffusion in palladium. *Phys. Rev. Lett.* **52**, 1625 (1984).
- Brouwer, R. C. & Griessen, R. Electromigration of hydrogen in alloys. Evidence of unscreened proton behavior. *Phys. Rev. Lett.* **62**, 1760 (1989).
- van der Sluis, P., Ouwerkerk, M. & Duine, P. A. Optical switches based on magnesium lanthanide alloy hydrides. *Appl. Phys. Lett.* **70**, 3356 (1997).

Acknowledgements. We thank B. Hjörvarsson, E. S. Kooij, M. W. J. Prins, P. van der Sluis, M. Ouwerkerk, P. H. L. Notten and P. J. Kelly for discussions. This work is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is supported by NWO; it was also supported in part by the European Commission through the TMR programme.

Correspondence and requests for materials should be addressed to S.J. v.d. M. (e-mail: sejan@nat.vu.nl)

Evidence for laser action driven by electrochemiluminescence

Tsutomu Horiuchi, Osamu Niwa & Noriyuki Hatakenaka

NTT Basic Research Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi, Kanagawa 243-0198, Japan

Emission of light from excited-state dye molecules can be driven by the electron transfer between electrochemically generated anion and cation radicals—a process known as electrochemiluminescence^{1–5} (ECL). ECL has been investigated for both display⁶ and laser applications^{7–9}. The latter is of particular interest as, in contrast to conventional dye lasers, a laser operating by this principle would not require an additional laser source optically to pump the dye into the required excited state, and may offer additional advantages in terms of power, tunability and range of available wavelengths. But the pumping rate hitherto achieved by ECL is two orders of magnitude lower than the optical pumping threshold⁹. Here we describe a device structure designed to enhance the efficiency of the ECL process, and present evidence that laser action has been realized in such a structure: the ECL spectrum is strongly modulated by the device structure, the output intensity shows a clear threshold as the drive current

increases, and spectral narrowing is observed as the intensity increases.

A diagram of this ECL experiment is shown in Fig. 1. A pair of platinum-film electrodes on quartz substrates were positioned facing each other 2–7 μm apart. These sputter-deposited platinum electrodes provided a mirrored surface. One platinum electrode was a half-mirror, prepared by controlling the deposition thickness. Its transmittance was 0.8%. 9,10-diphenylanthracene (DPA), which exhibits lasing with external optical pumping¹⁰, was dissolved in *N,N*-dimethylformamide (DMF) and the solution was continuously introduced into the cavity. The ground-state DPA molecules were oxidized and reduced at each electrode by applying a potential between the electrodes. Ground- and excited-state molecules were generated as the result of the annihilation reaction between anion and cation radicals, and the excited state emitted blue light (wavelength, 420 nm), which corresponded to the potential difference between the oxidation (1.3 V versus Ag/AgCl) and reduction (–1.7 V) potential of DPA. The excited-state DPA molecules returned to the initial state (ground-state molecules) and the same reaction continued as long as the molecules were within the cavity.

The structure shown in Fig. 1 is very useful for achieving highly efficient ECL with which to realize lasing. The d.c. electrolysis method generates anion and cation radicals more efficiently than the a.c. electrolysis method, because it is not affected by non-faradaic (charging) current. The smaller distance between the anode and cathode can greatly enhance the faradaic current as a result of redox cycling. As the very small interelectrode distance reduces resistance, we can generate ECL without adding a supporting electrolyte to act as a quencher. Moreover, the concentration of quencher generated by the side reaction did not increase when the electrolyte was made to flow. Furthermore, the cell had a Fabry–Pérot resonance structure which can be expected to increase the

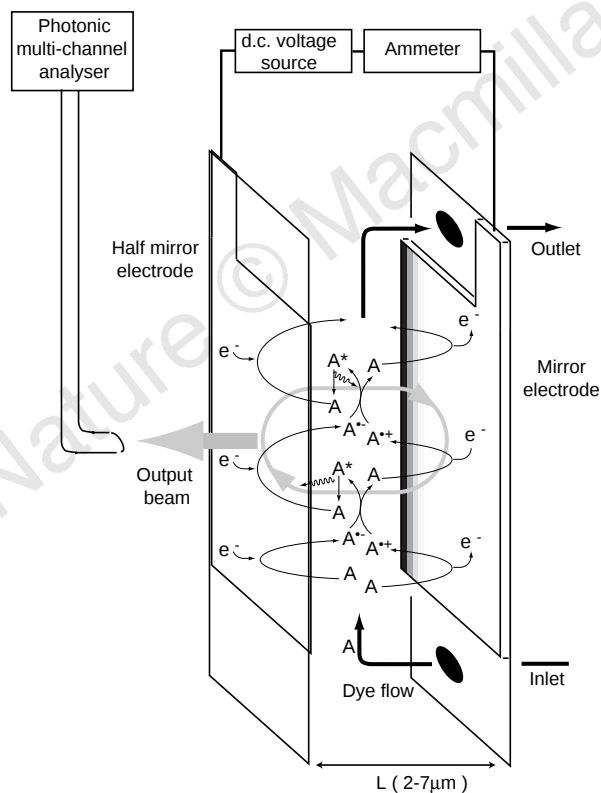


Figure 1 Diagram of electrochemiluminescence (ECL) laser. Mirror and half-mirror platinum electrodes ($2 \times 2 \text{ mm}$) on quartz substrates faced each other separated by a $4\text{-}\mu\text{m}$ -thick polypropylene spacer which had a rectangular hole. Dye solution flowed in the thin-layer cell formed by the two electrodes. To prevent leaks, the seam between the two substrates was filled with adhesive. ECL from the device was guided through an ultraviolet/visible multimode fibre (diameter of core and cladding was 400 and $430 \mu\text{m}$, respectively) and spectrum measurement was carried out using a photonic multi-channel analyser (CCD image sensor). A, A^* , $A^{\bullet-}$ and $A^{\bullet+}$ indicate respectively the ground, excited, anion radical and cation radical states of 9,10-diphenylanthracene (DPA).

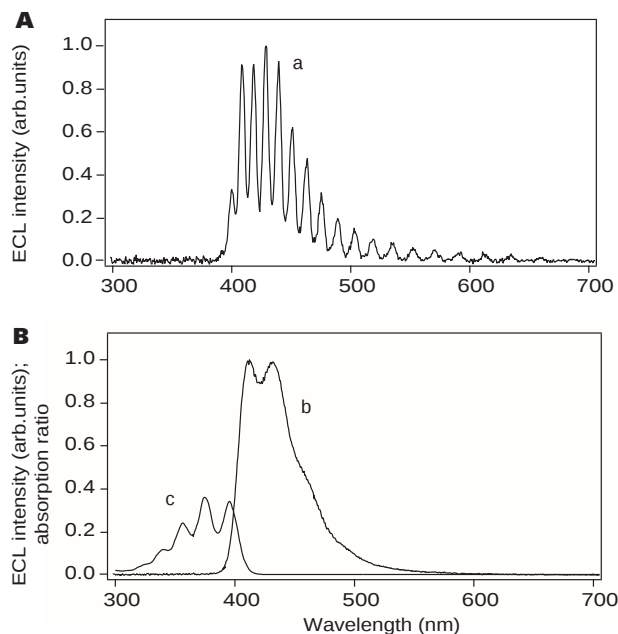


Figure 2 The effect of the thin-layer cell cavity on ECL. **A**, ECL spectrum of DPA from a thin-layer flow cell cavity. A solution of 25 mM DPA in DMF was introduced with a flow rate of $5 \mu\text{l min}^{-1}$. The DPA solution was electrolysed at 18 V using a d.c. voltage source. The exposure time of the photonic multi-channel analyser was 1,000 ms. **B**, Reference spectra. Trace b, the ECL spectrum of free space; trace c, the absorption spectrum of DPA. Both ECL spectra were normalized with their maximum values. The absorption spectrum was calculated to correspond to the actual cavity length ($6 \mu\text{m}$) and DPA concentration (25 mM), using data obtained from the absorption spectrum obtained with 10 mM and 0.1 mM DPA solutions in an optical cell with a 1-cm path length.

stimulated emission rate by effective use of the absorption/re-emission effect of photons from the short-wavelength region. In order to electrolyse the DPA at a high rate, a sufficiently high potential (above 10 V) was applied, taking account of the ohmic drop in the double-layer region.

Trace a in Fig. 2A shows the ECL spectrum observed from this thin-layer flow cell cavity, which was strongly modulated by Fabry–Pérot resonance. As a reference, the free-space ECL spectrum observed for a structure with no cavity is shown in trace b, Fig. 2B. Many clear resonant peaks were observed only when the thin-layer flow cell cavity was used. The distances between the peaks were in good agreement with the longitudinal mode separation ($\lambda^2/(2nL)$) determined by the cavity length (L) and refractive index (n). The actual cavity length (6 μm) was estimated as the sum of the polypropylene film (4 μm) and adhesive (2 μm) layer thickness. (See Fig. 1 legend for construction details.) As shown in Fig. 2B, trace c, the ECL and absorption spectra overlap, and $\sim 27\%$ of the photons at the 400-nm wavelength were absorbed in the DPA solution. In contrast, very few photons were absorbed in the long-wavelength region (0.0049% at 500 nm).

The ECL intensity increases nonlinearly with increasing faradaic current through the cell (injection current), as shown in Fig. 3. The current was controlled by changing the voltage. Both current and intensity (peak height at 445 nm) were measured in a steady state. Theoretically, ECL intensity is proportional to the square of the DPA concentration which is proportional to the faradaic current. Therefore, the ECL intensity was expected to be proportional to the square of the current. However, the observed intensity could not be represented as a quadratic function of the current. A clear threshold was observed at 0.45 mA, leading to similarity for the intensity–current curve of an electrically pumped semiconductor laser.

This threshold current is roughly estimated as $m\epsilon\gamma/(\phi\tau g)$ from rate equations: here m is the number of transverse modes in the cavity, which is the ratio of electrode area (4 mm²) to the effective area (3.52 μm^2) occupied by a single mode resulting from diffraction loss in the cavity¹¹; ϕ and τ are respectively the ECL quantum yield of 0.08 (ref. 3) and the lifetime of the excited DPA (5×10^{-9} s; ref. 7); γ and g are respectively the cavity decay rate and gain rate in cm^{-1} ; and e is the electron charge. The γ value of 1,089 cm^{-1} was obtained from the cavity length (7.2 μm) and the reflectivity of Pt, the g value of 1,205 cm^{-1} was obtained from γ , and the net gain of -116 cm^{-1} at 445 nm was derived by the Hakki–Paoli method^{12,13} which involved analysing adjacent peak maxima and the intermedi-

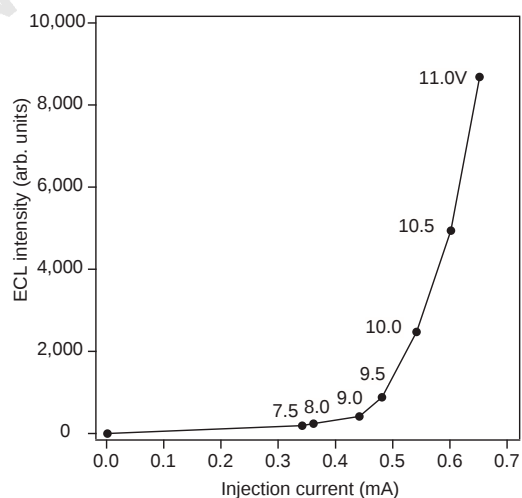


Figure 3 ECL intensity (peak height at 445 nm) versus injection current. A solution of 3 mM DPA in DMF was introduced with a flow rate of 5 $\mu\text{l min}^{-1}$. The exposure time was 1,000 ms. The actual cavity length was 7.2 μm . The labels near the points are the applied potentials. A clear threshold was observed at 0.45 mA.

ate valley of the ECL spectrum. The Pt reflectivity used in the calculation was the actual value of 0.22 (at 445 nm) obtained from the linewidth. The threshold current of 0.41 mA was obtained using the above values and was of the same order as the experimental value, 0.45 mA.

In Fig. 4, we compare two spectra, one above and one well below the threshold, to confirm the possibility of lasing. We do this because the spectrum above the threshold should be obviously different from that below the threshold. The peak position and distance were almost the same for the two spectra. However, the peak height of the strong emission ECL in the short-wavelength region was much higher than that of the weak one, and the envelopes of the modulated spectrum narrowed. This envelope-narrowing effect is similar to that observed in ordinary laser devices below and near the lasing threshold. We also plot gain/loss spectra in the inset of Fig. 4, derived from ECL spectra by the Hakki–Paoli method^{12,13}. The negative value was obtained only in the high-intensity ECL emission spectrum (trace a in the inset), which suggests that the gain surpasses the absorption loss. The high gain in the 440–500 nm wavelength region was caused by the low absorption in the DPA solution and the absorption/re-emission of photons from the short-wavelength region.

A much clearer spectrum profile change would be observed if a much stronger emission were to be realized. This could be achieved by reducing the cavity length, and increasing the electrochemical reaction efficiency with a large potential and high flow rate. Figure 5 shows the time-dependent change in the spectrum after applying a sufficiently high voltage (30 V). The spectra were sequentially measured with a 50-ms exposure time. As the diffusion of the molecules to the electrode is a rate-determining step, the ECL is very

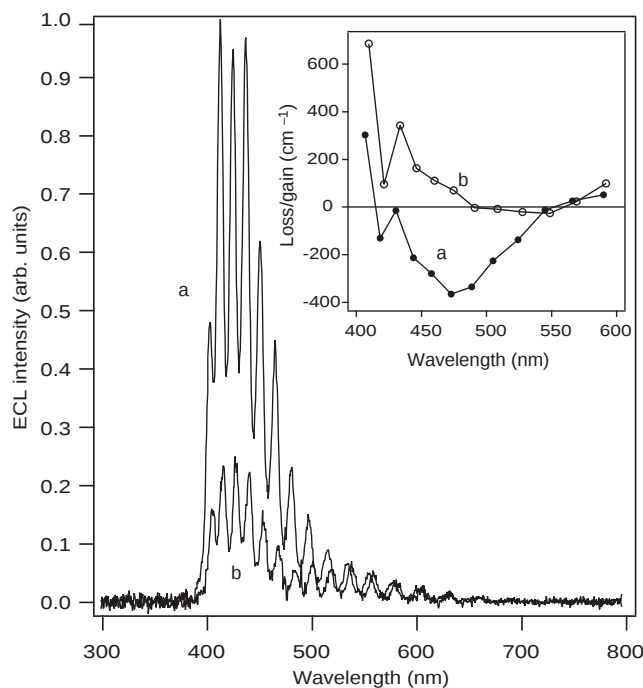


Figure 4 A comparison of the ECL spectrum above the threshold (trace a), with that well below the threshold (trace b). A solution of 20 mM DPA in DMF was introduced at a flow rate of 10 $\mu\text{l min}^{-1}$. 20 V was applied and the exposure time was 20 ms (trace a); 4 V was applied and a tenfold integrated experiment was undertaken with a 20,000 ms exposure time (trace b). Normalization was carried out in order to adjust the peak height around 550 nm. The peak height of trace a around 550 nm was 416 times larger than that of trace b for the same exposure time. The actual cavity length was 5.1 μm . Inset, gain/loss spectra calculated from corresponding spectra. The reflectivity of Pt used for the calculation was 40% of the ideal value. The labels correspond to those of the ECL spectra.

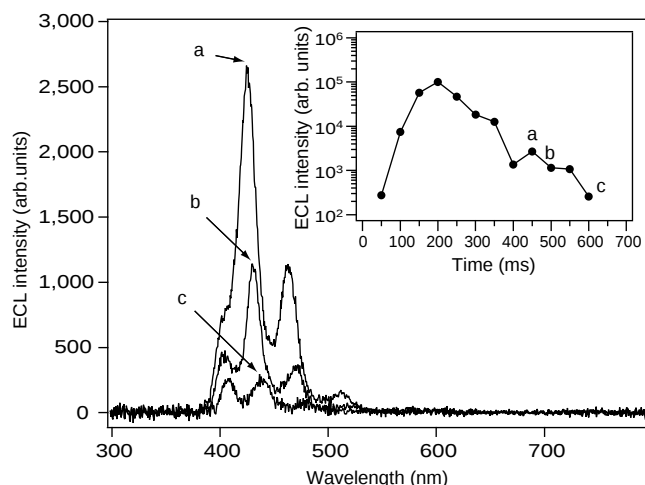


Figure 5 The change of ECL spectrum with time after application of 30 V to the cell. A solution of 25 mM DPA in DMF was introduced with a flow rate of $20 \mu\text{L min}^{-1}$. The spectra were sequentially measured with a 50-ms exposure time from just before the application of 30 V. The actual cavity length was $1.9 \mu\text{m}$. Inset, the time decay of the peak height located at $\sim 435 \text{ nm}$. The time origin is the last time at which the ECL signal was within the noise level. The labels correspond to those on the ECL spectra.

bright in the initial stage and then its intensity gradually decreases. This is because the concentration of the ground-state DPA is very high near the electrode and generates a higher concentration of excited-state DPA than that in the steady state. The temporal change in peak intensity at 435 nm is shown in the inset. The decay time was 0.071 s. This long decay, resulting from the continuous supply of dye molecules, depended on the flow rate and applied voltage.

In the low-intensity spectrum (trace c in Fig. 5), the peaks at 405 and 435 nm are almost the same height. However, in the high-intensity spectrum (trace b in Fig. 5), the peak at 435 nm is greatly amplified compared with the peak at 405 nm. The linewidth (full-width at half-maximum) of the peak at 435 nm decreased $\sim 25\%$. The peak at 405 nm merged with the peak at 435 nm when the intensity increased further (trace a in Fig. 5). The merged peak still shows line narrowing with increasing ECL intensity. For example, the FWHM of the peak at the highest intensity (at 200-ms in the inset) is $\sim 32\%$ narrower than that at lower intensity (300 ms). These spectrum changes as a function of intensity clearly indicate a stimulated emission.

The results shown in Figs 3–5 indicate that the lasing was achieved by an electrochemical reaction. This is supported by another experiment in which fringe patterns were observed when the half-mirror was replaced with an interdigitated array electrode which functioned as a grating. Much clearer lasing could be observed by either decreasing the distance between the electrodes of the flow cell to the order of a wavelength, or increasing their reflectivity. □

Received 9 June 1997; accepted 9 June 1998.

1. Feldberg, S. W. Theory of controlled potential electrogeneration of chemiluminescence. *J. Am. Chem. Soc.* **88**, 390–393 (1966).
2. Maloy, J. T., Prater, K. B. & Bard, A. J. Electrogenerated chemiluminescence. II. The rotating ring-disk electrode and the pyrene- N,N,N',N' -tetramethyl- p -phenylenediamine system. *J. Phys. Chem.* **72**, 4348–4350 (1968).
3. Keszthelyi, C. P., Tokel-Takvoryan, N. E. & Bard, A. J. Electrogenerated chemiluminescence: Determination of the absolute luminescence efficiency in electrogenerated chemiluminescence. 9,10-Diphenylanthracene-thianthrene and other systems. *Anal. Chem.* **47**, 249–256 (1975).
4. Bartelt, J. E., Drew, S. M. & Wightman, R. M. Electrochemiluminescence at band array electrodes. *J. Electrochem. Soc.* **139**, 70–74 (1992).
5. Collinson, M. M., Pastore, P., Maness, K. M. & Wightman, R. M. Electrochemiluminescence interferometry at microelectrodes. *J. Am. Chem. Soc.* **116**, 4095–4096 (1994).
6. Schaper, H., Koestlin, H. & Schneider, E. New aspects of d.c. electrochemiluminescence. *J. Electrochem. Soc.* **129**, 1289–1294 (1982).
7. Measures, R. M. Prospects for developing a laser based on electrochemiluminescence. *Appl. Opt.* **13**, 1121–1133 (1974).
8. Measures, R. M. Physical constraints associated with the development of a laser based on electrochemiluminescence. *Appl. Opt.* **14**, 909–916 (1975).

9. Heller, C. A. & Jernigan, J. L. Electrochemical pumping of laser dyes. *Appl. Opt.* **16**, 61–66 (1977).
10. Myer, J. A., Johnson, C. L., Kierstead, E., Sharma, R. D. & Itzkan, I. Dye laser stimulation with a pulsed N_2 laser line at 3371 Å. *Appl. Phys. Lett.* **16**, 3–5 (1970).
11. Ujihara, K. A simple relationship between directivity and linewidth of a planar microcavity laser. *Jpn. J. Appl. Phys.* **33**, 1059–1060 (1994).
12. Hakki, B. W. & Paoli, T. L. CW degradation at 300°K of GaAs double-heterostructure junction lasers. II. Electronic gain. *J. Appl. Phys.* **44**, 4113–4119 (1973).
13. Hakki, B. W. & Paoli, T. L. Gain spectra in GaAs double-heterostructure injection lasers. *J. Appl. Phys.* **46**, 1299–1306 (1975).

Acknowledgements. We are grateful to M. Morita for encouragement and support.

Correspondence and requests for materials should be addressed to T.H. (e-mail: horiuchi@will.brl.nitt.co.jp).

Effects of orbital decay on satellite-derived lower-tropospheric temperature trends

Frank J. Wentz & Matthias Schabel

Remote Sensing Systems, 438 First Street, Suite 200, Santa Rosa, California 95401, USA

The 17-year lower-tropospheric temperature record derived from the satellite Microwave Sounding Unit (MSU)^{1–3} shows a global cooling trend, from 1979 to 1995, of $-0.05 \text{ K per decade}$ at an altitude of about 3.5 km (refs 4, 5). Air temperatures measured at the Earth's surface, in contrast, have risen by approximately $+0.13 \text{ K per decade}$ over the same period^{4,6}. The two temperature records are derived from measurements of different physical parameters, and thus are not directly comparable. In fact, the lower stratosphere is cooling substantially (by about $-0.5 \text{ K per decade}$)⁵, so the warming trend seen at the surface is expected to diminish with altitude and change into a cooling trend at some point in the troposphere. Even so, it has been suggested that the cooling trend seen in the satellite data is excessive^{4,7,8}. The difficulty in reconciling the information from these different sources has sparked a debate in the climate community about possible instrumental problems and the existence of global warming^{4,7,9}. Here we identify an artificial cooling trend in the satellite-derived temperature series caused by previously neglected orbital-decay effects. We find a new, corrected estimate of $+0.07 \text{ K per decade}$ for the MSU-based temperature trend, which is in closer agreement with surface temperatures. We also find that the reported⁷ cooling of the lower troposphere, relative to the middle troposphere, is another artefact caused by uncorrected orbital-decay effects.

The debate within the climate community^{4,7,9} about differences between the global temperature trends determined from *in situ* surface records and from lower-tropospheric MSU satellite data has centred on two fundamental issues: physical differences between the two measurements and possible instrumental artefacts. MSU middle tropospheric temperature measurements (MSU2) represent a vertically weighted air temperature centred at an altitude of $\sim 7 \text{ km}$, while the MSU lower-tropospheric temperature (MSU2R) weighting function peaks around 3.5 km. MSU2, MSU2R and *in situ* surface temperatures are therefore not directly comparable, and their decadal trends are thus not expected to be identical, particularly as the lower stratosphere is known to be cooling substantially⁵. However, the large difference between global surface and MSU2R temperature trends ($+0.13 \text{ K per decade}$ and $-0.05 \text{ K per decade}$, respectively^{4–6}) is difficult to explain. In the tropics the difference is even more pronounced, with sea surface temperatures rising by $+0.10 \text{ K per decade}$ and MSU2R falling by $-0.11 \text{ K per decade}$ ⁷. The

magnitude of these differences has led to the suggestion^{4,7,8} that the reported MSU2R cooling trend, particularly in the tropics, is excessive.

In addition to the concern over excessive cooling of MSU2R compared to *in situ* data, there is another more obvious problem: MSU2R, centred at around 3.5 km altitude, is cooling at -0.17 K per decade relative to MSU2, centred at around 7 km (ref. 7). This strong relative cooling disagrees with radiosonde data that show little relative difference in the temperature trends observed at 850, 700 and 500 mbar levels¹⁰ (corresponding to altitudes of 1.4, 3.0 and 5.6 km). Recent radiosonde results derived for the MSU time period confirm this observation: the temperature trends of -0.01 , -0.07 and -0.05 K per decade for 850, 700 and 500 mbar levels, respectively (D. Parker, personal communication) are fairly similar. These values indicate that the lower troposphere (850–700 mbar) is slightly warming relative to the middle troposphere (700–500 mbar), which disagrees with the very large cooling of MSU2R relative to MSU2. The discrepancy between the relative trends in the satellite and radiosonde data is particularly difficult to explain, considering that 20% of the MSU2R signal comes from the surface (discussed below), which is warming, while a substantial fraction of the MSU2 signal comes from the lower stratosphere, which is cooling. This seemingly unphysical cooling of MSU2R relative to MSU2 led us to examine closely the derivation of the two parameters, thereby uncovering a subtle effect related to orbital decay that significantly affects MSU2R but not MSU2.

MSU tropospheric temperatures are derived from measurements of Earth's radiance at 53.74 GHz (MSU channel 2). The MSU sensor scans perpendicular to the satellite subtrack, with 11 footprints extending over a range of incidence angles from 0° to 55° . The MSU2 middle-tropospheric temperature is found from a limb-corrected average of the central 9 footprints. MSU2 has a broad

vertical weighting function centred at ~ 7 km but retaining appreciable weight from the surface well into the stratosphere. To retrieve an air temperature that is more indicative of the lower troposphere, a deconvolution is performed, making use of the fact that the weighting function for the near-limb observations (44° to 55°) peaks somewhat higher in altitude than for the near-nadir observations (22° to 33°). The MSU2R lower-troposphere temperature T_{2R} is calculated from the following equation²:

$$T_{2R} = 4T_N - 3T_L \quad (1)$$

where T_N and T_L are the near-nadir and near-limb observations, respectively. In principle, this linear combination of T_N and T_L leads to an effective weighting function centred near 3.5 km.

Our investigation reveals that the unphysical cooling of MSU2R relative to MSU2 is a direct result of T_L warming relative to T_N . In the tropics from 1979 to 1995, T_N and T_L increased by $+0.05$ K per decade and $+0.10$ K per decade, respectively, resulting in an MSU2R trend of -0.11 K per decade. Thus, although T_N and T_L are both warming, MSU2R exhibits a cooling trend arising from the extrapolation to the lower troposphere. The question then becomes: why is T_L warming by $+0.05$ K per decade relative to T_N ?

All of the NOAA polar orbiting platforms carrying the MSU instruments lose altitude after launch. In constructing the MSU time series, the satellites are sequentially intercalibrated, so the effect of altitude decrease of one satellite is added onto the next. Figure 1 shows the altitude changes of NOAA-6 to NOAA-14 referenced to a common baseline (NOAA-8 is excluded because it was not used in the MSU time series). Absolute orbital altitudes found from two-line orbital element data provided by NORAD are plotted in Fig. 1 inset. Immediately apparent in the figure is the presence of two intervals, 1979–83 and 1989–92, with dramatically steeper slopes. These intervals correspond to periods during which solar activity was at a maximum¹¹. Higher levels of solar ultraviolet radiation during these periods heat the upper atmosphere and increase the drag on the spacecraft. Over intervals of high solar activity, the observed orbital decay rate is ~ 2 km yr^{-1} , while periods of low solar activity show a smaller rate of 0.3 km yr^{-1} . Over the entire 17-year record there is a net drop of 20 km, equivalent to an average rate of 1.2 km yr^{-1} .

This decrease in altitude is significant because it leads to a corresponding decrease in the Earth incidence angle due to the sphericity of the Earth's surface. For the near-nadir observations, the incidence angle decreases by $0.0049^\circ \text{yr}^{-1}$ on average, while the near-limb angle changes by more than twice as much ($0.0120^\circ \text{yr}^{-1}$). Radiative transfer calculations¹² show that these changes in incidence angle increase T_N and T_L by $+0.008$ K per decade and $+0.052$ K per decade, respectively. Thus the orbital decay causes T_L to warm relative to T_N at a rate of 0.044 K per decade, in close agreement with the anomalous trend mentioned above. Calculations were performed for several atmospheric profiles (high latitudes, mid-latitudes and tropics) and for both land and ocean surfaces. The land emissivity was fixed at 0.85, and the ocean emissivity comes from a microwave ocean model¹². Atmospheric absorption due to oxygen and water vapour is based on a standard microwave propagation model¹³. The orbital decay correction is insensitive (± 0.007 K per decade) to latitude, surface type, and the details of the radiative transfer model. Considering the error in specifying the altitude decay and the small variability due to surface type and latitude, the error in specifying the orbital decay correction is ± 0.01 K per decade at a 95% confidence level.

Removing the effect of orbital decay, we find that the corrected T_N and T_L are both warming at nearly the same rate. In view of their very broad weighting functions, with centroids differing by only 1 km, it seems quite reasonable that the two should have similar trends. Substituting the corrected values of T_N and T_L in equation (1) results in a $+0.12$ K per decade global correction to MSU2R for the 1979–95 period. Adding this correction to the previously

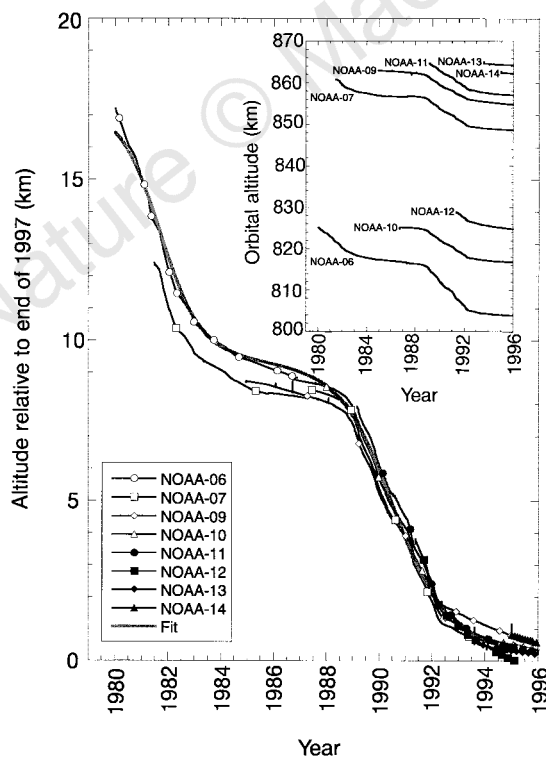


Figure 1 Cumulative orbital decays for NOAA satellites from 1979 to 1996. Data for all satellites used in the MSU lower-tropospheric temperature series are plotted, shifted by constant offsets to correct for differences in orbital altitude. An empirical nonlinear regression to the data is shown by the grey curve. Absolute altitudes are shown in the inset. (Data shown in this figure are available at <http://www.grove.net/~tkelso/NORAD/archives>)

reported trends (-0.11 K per decade in the tropics and -0.05 K per decade globally) results in new trends of $+0.01$ K per decade in the tropics and $+0.07$ K per decade globally.

The decay correction varies from year to year, being greatest during periods of maximum solar activity. Figure 2 shows the effect of orbit decay on trends in MSU2R tropical average (20°S – 20°N) anomaly time series⁷. Horizontal bars at the top of the figure indicate the periods of high solar activity. To calculate the decay correction, plotted in Fig. 2c, the altitude data for NOAA-6 and NOAA-10, which taken together span the entire 17-year period, were merged at the beginning of 1987. The resulting curve was regressed to an empirical function (the broad grey line in Fig. 1) and used to compute the geometrical correction to MSU2R. Figure 2a shows the time series for the uncorrected MSU2R temperature anomalies (dashed), with its downward trend of -0.11 K per decade. The corrected MSU2R time series is plotted as the solid line. Figure 2b shows the difference between MSU2R and MSU2 temperature anomalies, with and without the correction for orbital decay (solid and dashed lines, respectively). MSU2 is much less sensitive to the effect of orbit decay, serving as a reference for evaluating this effect on MSU2R. Clearly evident in this figure is the cooling of the uncorrected MSU2R relative to MSU2 that occurs during the two periods of high solar activity, and correspondingly large orbital decay. The correlation coefficient between the uncorrected MSU2R–MSU2 difference and the decrease in satellite altitude is 0.91. To compute this correlation, the MSU2R–MSU difference is first smoothed with a gaussian smoothing function with a 180-day standard deviation. The high correlation is strong evidence that the cooling of MSU2R during 1979–83 and 1989–92 is due to orbital decay.

Although orbital decay affects other satellite instruments, it is the unavoidable noise amplification associated with the MSU2R deconvolution procedure and the very small decadal trends being sought which makes this particular data set so sensitive. Other satellite-measured parameters will probably be much less affected. For example, the net spurious warming introduced into MSU2 is only $+0.008$ K per decade. Nevertheless, the effect is easily computed and should be accounted for in future analyses of satellite data.

Despite the much improved consistency between MSU2 and MSU2R and the closer agreement of the MSU2R record with surface temperature data, there remain some questions which are not entirely resolved by our correction. The main issue is the reported⁹ good agreement between the MSU2R (before correction) and radiosonde measurements. Both data sets implied a lower-tropospheric temperature trend of about -0.05 K per decade. However,

these comparisons did not account for surface effects. MSU2R is not simply a measure of the lower-troposphere temperature, but also contains an Earth surface emission component. This sensitivity of MSU2R to surface emissions results from the factor of 4 applied to T_N in equation (1). We use the radiative transfer model to simulate this sensitivity. The model assumes the Earth's surface temperature changes by 1 K while the temperature at an altitude of 1.5 km (850 mbar) remains constant. Between the surface and 850 mbar, the change linearly decreases to 0 K. The increase of 1 K in the near-surface temperature produces increases of 0.25 K and 0.17 K in the MSU2R-derived temperature (T_{2R}) for land and ocean surfaces, respectively, even though the lower-tropospheric temperature remains constant. Thus, on average, $\sim 20\%$ of the observed trends of T_{2R} will be due to near-surface trends, and the remainder will be due to trends in the lower and middle troposphere (850–300 mbar). The MSU2 product, which approximately equals T_N , is much less sensitive to surface effects because of the absence of the factor of 4 in equation (1). Therefore, to directly compare MSU2R with radiosondes, a surface temperature layer is added to the radiosonde layers, and a vertical integration over all layers is done to compute an effective MSU2R trend of -0.02 K per decade (D. Parker, personal communication), which is in closer agreement with the observed $+0.07$ K per decade trend.

In interpreting these results it is important to consider the uncertainties in estimating decadal trends from MSU and radiosondes. The error in the MSU2R trend is estimated to be ± 0.05 K per decade at a confidence level of 80% (J. Christy, personal communication). This error was determined by comparing the results of using different combinations of satellites and by reconstructing the time series with random variations of inter-satellite biases. There is a good deal of subjectivity in the MSU error analysis, and, as we show here, there is the possibility that the true error bar is larger due to other, yet to be discovered, effects. The MSU error estimate does not include the ± 0.01 K per decade uncertainty in specifying the orbit decay correction. This uncertainty must be added to the overall error bar. The error in the radiosonde trend is ± 0.10 K per decade at a 95% confidence level (D. Parker, personal communication). In this case, the error bars are based on standard least-squares theory but take account of the autocorrelation of the residuals from the linear trend. The error estimates do not allow for systematic biases in the data or for the geographical sparsity of sampling¹⁴. So again, there is the possibility that the true error is larger. Thus, considering the size and uncertainty of the error bars, the disagreement of 0.09 K per decade between MSU2R and radiosondes may not be too significant.

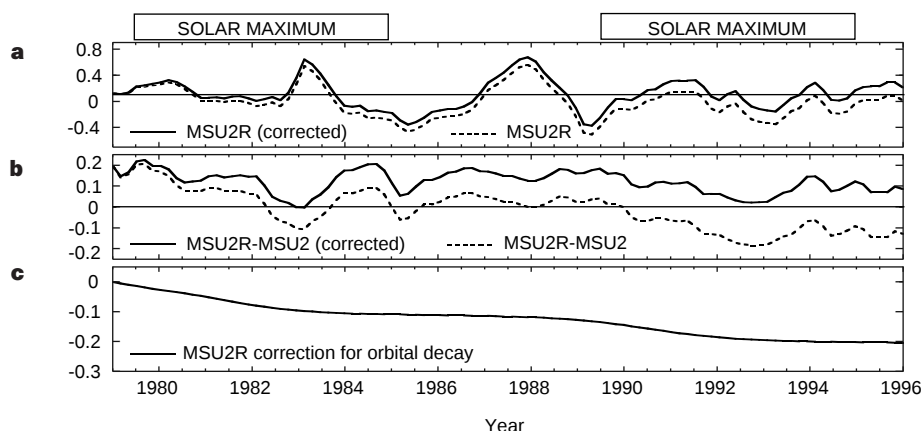


Figure 2 Effect of correction for orbital decay on MSU2R lower-tropospheric temperature anomalies. **a**, Time series for tropical average (20°S – 20°N) MSU2R lower-tropospheric temperature anomalies with (solid line) and without (dashed line) the correction for orbital decay. **b**, Differences between the lower-(MSU2R)

and mid-(MSU2) tropospheric temperatures, plotted as in **a**, **c**. The orbital correction to MSU2R based on the data in Fig. 1 (data in this figure adapted from Fig. 1 of ref. 7).

The orbit decay correction of +0.12 K per decade is a relatively simple, though easily overlooked, geometric effect resulting from decreasing satellite altitude. This correction eliminates the spurious cooling of MSU2R during 1979–83 and 1989–92 as well as the spurious cooling of the lower troposphere relative to the middle troposphere. The correction also leads to an overall global warming trend of +0.07 K per decade for the 1979–95 period, which is in closer agreement with surface temperatures. However, considering the size and uncertainty of the MSU2R error bars, we advise caution with respect to the significance placed on the absolute trend of +0.07 K per decade. The derivation of trends from multiple satellites is a complex process, particularly considering the fact that MSU2R is an extrapolation from the middle to lower troposphere: subtle, unmodelled effects can significantly alter the trend estimate. The MSU data set needs to be more closely examined, and a more rigorous error analysis should be done.

Although the MSU and radiosondes have provided an extremely valuable air-temperature record for the past few decades, they were originally designed with meteorological objectives in mind, and their ability to measure very small climate variations is limited. There is a need for future climate-monitoring systems that are more accurate and robust. □

Received 24 February; accepted 7 July 1998.

1. Spencer, R. W. & Christy, J. R. Precise monitoring of global temperature trends from satellites. *Science* **247**, 1558–1562 (1990).
2. Spencer, R. W. & Christy, J. R. Precision and radiosonde validation of satellite gridpoint temperature anomalies. Part II: A tropospheric retrieval and trends during 1979–90. *J. Clim.* **5**, 858–866 (1992).
3. Christy, J. R. & McNider, R. T. Satellite greenhouse signal. *Nature* **367**, 325 (1994).
4. Hurrell, J. W. & Trenberth, K. E. Difficulties in obtaining reliable temperature trends: Reconciling the surface and satellite MSU2R trends. *J. Clim.* (in the press).
5. Houghton, J. T. *et al.* (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).
6. Jones, P. D. Recent warming in global temperature series. *Geophys. Res. Lett.* **21**, 1149–1152 (1994).
7. Hurrell, J. W. & Trenberth, K. E. Spurious trends in satellite MSU temperatures from merging different satellite records. *Nature* **386**, 164–167 (1997).
8. Hansen, J. *et al.* Satellite and surface temperature data at odds? *Clim. Change* **30**, 103–117 (1995).
9. Christy, J. R., Spencer, R. W. & Braswell, W. D. How accurate are satellite 'thermometers'? *Nature* **389**, 342–342 (1997).
10. Hansen, J. *et al.* Forcings and chaos in interannual to decadal climate change. *J. Geophys. Res.* **102**, 25679–25720 (1997).
11. Willson, R. C. Total solar irradiance trend during solar cycles 21 and 22. *Science* **277**, 1963–1965 (1997).
12. Wentz, F. J. A well-calibrated ocean algorithm for special sensor microwave/imager. *J. Geophys. Res.* **102**, 8703–8718 (1997).
13. Liebe, H. J. An updated model for millimeter wave propagation in moist air. *Radio Sci.* **20**, 1069–1089 (1985).
14. Gaffen, D. Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res.* **99**, 3667–3676 (1994).

Correspondence and requests for materials should be addressed to F.J.W. (e-mail: wentz@remss.com).

Isotopic evidence for a solar argon component in the Earth's mantle

R. O. Pepin

School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA

Determining the presence of solar argon, krypton and xenon in the Earth's mantle is important for understanding the source, incorporation mechanism and transport of noble gases in the Earth, as well as the evolutionary history of the Earth's atmosphere. There are strong indications in the mid-ocean ridge basalt database that solar helium and neon are indeed present^{1–3}, and modelling exercises indicate that the compositions of all five noble gases in the Earth's primordial inventory were solar-like^{3–5}. But solar isotopic signatures of the heavier noble gases argon and xenon, which differ significantly from atmospheric compositions,

have appeared only subtly if at all in analyses of mantle-derived samples⁶—their non-radiogenic isotope ratios are generally found to be indistinguishable or only slightly different from those in the atmosphere^{7–10}. The first promising isotopic evidence for a solar-like argon component in the Earth's mantle appeared in a recent analysis of basalt glasses from the Hawaiian Loihi seamount¹¹. Here I show that recent measurements¹² of neon and argon isotopes in a suite of mid-ocean ridge basalt samples from the southern East Pacific Rise greatly strengthen the case for the presence of solar argon, and by inference krypton and xenon, in the Earth's mantle.

Three recent publications bear directly on the issue of a solar component in ocean basalt source reservoirs. Burnard *et al.*¹³ used data from a volatile-rich mid-ocean ridge basalt (MORB) glass ('popping rock' 2πD43, so named for the audible pops of exploding gas-filled vesicles immediately after it was dredged from the deep Mid-Atlantic Ridge) together with modelling arguments to infer a qualitatively solar-like elemental abundance pattern for all five noble gases in the mantle. They suggested that observation of ³⁸Ar/³⁶Ar ratios between the atmospheric value of 0.1880 and the lower solar ratio would confirm their proposition of the presence of solar Ar, but were pessimistic about the chances for resolving this signature from pervasive atmospheric contamination. Soon thereafter, however, Valbracht *et al.*¹¹ reported just such a finding in Loihi glasses, with ³⁸Ar/³⁶Ar as low as 0.1848 ± 0.0012 in one stepwise degassing fraction, and further showed that sample-to-sample variations in ³⁸Ar/³⁶Ar and ²⁰Ne/²²Ne were correlated in ways highly suggestive of mixing between atmospheric and solar-like endpoint compositions. The third publication, by Niedermann *et al.*¹² on He, Ne and Ar distributions in MORB glasses from the East Pacific Rise (EPR), appeared shortly before the Loihi study. These EPR samples also have low ³⁸Ar/³⁶Ar ratios, here down to 0.1845 ± 0.0018 (1σ) in a total extraction and 0.1814 ± 0.0019 in a thermal release fraction from the same sample (G184D). Niedermann and co-workers noted the presence of these sub-atmospheric values in their data set¹² but did not discuss them further. I have examined the EPR data on a ³⁸Ar/³⁶Ar versus ²⁰Ne/²²Ne diagram, and find that they define a remarkably coherent mixing curve between air-derived Ne and Ar and a second component isotopically consistent with current estimates for the solar wind.

The Ne and Ar compositions obtained by Niedermann *et al.*¹² are listed in Table 1 and plotted against each other in Fig. 1. The solar-wind ³⁸Ar/³⁶Ar range of ~0.1786–0.1724 is based on recent analyses^{14,15} which indicate substantially lower values for this ratio than an earlier estimate¹⁶, also shown, of 0.1825. In the following two-component modelling of the Fig. 1 distribution, I take the non-atmospheric component to be solar-wind Ne and Ar within their isotopic uncertainty ranges, and call this 'solar' with the assumption that the wind correctly represents the composition of the source supplying these gases to the early Earth. Isotope ratios along an air ⇌ solar (a ⇌ ☉) mixing trajectory may be shown to be

$$(^{20}\text{Ne}/^{22}\text{Ne})_{\text{calc}} = \left[\frac{(^{20}\text{Ne}/^{22}\text{Ne})_{\text{a}} + R(^{20}\text{Ne}/^{22}\text{Ne})_{\text{☉}}}{1 + R} \right] \quad (1)$$

$$(^{38}\text{Ar}/^{36}\text{Ar})_{\text{calc}} =$$

$$\left[\frac{[(^{36}\text{Ar}/^{22}\text{Ne})_{\text{a}} F_{36}^{\text{a}}](^{38}\text{Ar}/^{36}\text{Ar})_{\text{a}} + R[(^{36}\text{Ar}/^{22}\text{Ne})_{\text{☉}} F_{36}^{\text{☉}}](^{38}\text{Ar}/^{36}\text{Ar})_{\text{☉}}}{[(^{36}\text{Ar}/^{22}\text{Ne})_{\text{a}} F_{36}^{\text{a}}] + R[(^{36}\text{Ar}/^{22}\text{Ne})_{\text{☉}} F_{36}^{\text{☉}}]} \right] \quad (2)$$

where $R = [^{22}\text{Ne}]_{\text{☉}}/[^{22}\text{Ne}]_{\text{a}}$ is the Ne mixing ratio of the two components. F_{36}^{a} and $F_{36}^{\text{☉}}$ are defined and discussed below.

Elemental and isotopic ratios for the two end-point compositions in these equations are listed in Table 1. The factor $F_{36}^{\text{☉}} = (^{36}\text{Ar}/^{22}\text{Ne})_{\text{mantle}}/(^{36}\text{Ar}/^{22}\text{Ne})_{\text{☉}}$ is included to represent possible elemental fractionation of the solar component before, or accompanying, initial incorporation into its primordial mantle reservoir, or

Table 1 Isotopic data for the EPR MORB samples, terrestrial air and the solar wind

MORB sample	^{22}Ne (10^{-12} cc STP g $^{-1}$)	^{36}Ar (10^{-11} cc STP g $^{-1}$)	$^{20}\text{Ne}/^{22}\text{Ne}$	$^{38}\text{Ar}/^{36}\text{Ar}$	$^{40}\text{Ar}/^{36}\text{Ar}$
Total extractions					
G128G	33.2±0.8	20.4±1.5	10.47±0.06	0.1861±0.0018	3,770±260
G154D	87.8±1.9	32.0±1.5	10.20±0.04	0.1871±0.0011	4,370±190
G178D	12.5±0.5	24.8±2.7	11.61±0.07	0.1867±0.0014	870 $^{+130}_{-55}$
G184D	10.7±0.2	10.6±0.8	11.96±0.06	0.1845±0.0018	2,640 $^{+280}_{-70}$
G193G	20.1±0.4	13.8±1.0	11.21±0.05	0.1882±0.0014	1,510±110
G287G	25.3±0.5	39.4±2.0	11.25±0.04	0.1859±0.0007	940±45
G296G	26.5±0.5	7.96±0.59	11.02±0.07	0.1851±0.0021	3,340 $^{+340}_{-130}$
G296G*	15.2±1.2	85.9±9.1	12.08±0.21	0.1817±0.0014	580±55
G324G-1	21.0±1.2	17.8±1.0	10.71±0.06	0.1869±0.0010	2,660±140
G324G-2	42.1±0.9	48.0±1.3	10.25±0.04	0.1873±0.0006	1,144±15
G343G	6.67±0.42	10.1±0.6	11.09±0.19	0.1852±0.0019	1,330±65
Stepwise release fractions					
G128*, 1950 °C	2.27±0.08	0.94±0.23	12.49±0.18	0.1830±0.0055	13,200±3,250
G184D, 800 °C	5.77±0.16	1.53±0.09	12.71±0.07	0.1814±0.0019	5,280±250
G296G*, ≥1,600 °C	10.9±0.3	86.0±7.4	12.81±0.13	0.1817±0.0014	577±60
G343G*, 800 + 1,400 °C	3.21±0.13	1.87±0.30	11.60±0.20	0.1852±0.0024	1,670±360
Isotopic and elemental ratios					
$(^{20}\text{Ne}/^{22}\text{Ne})_{\text{a}}$	$(^{20}\text{Ne}/^{22}\text{Ne})_{\odot}$	$(^{38}\text{Ar}/^{36}\text{Ar})_{\text{a}}$	$(^{38}\text{Ar}/^{36}\text{Ar})_{\odot}$	$[^{36}\text{Ar}/^{22}\text{Ne}]_{\text{a}}$	$[^{36}\text{Ar}/^{22}\text{Ne}]_{\odot}$
9.80±0.08 (ref. 26)	13.80±0.10 (ref. 16)	0.1880±0.0005 (ref. 27)	0.1786–0.1724 (ref. 14) (ref. 15)	18.73±0.14 (ref. 28)	0.294±0.019 (ref. 29)

The top part of this table shows measured Ne and Ar concentrations in EPR basalt glasses (data from Tables 1 and 2 in ref. 12). The bottom part of the table shows isotopic and elemental ratios in air (a subscript) and the solar wind (⊙) (data from sources shown). Errors are 1σ.
* "Incompletely degassed" samples¹².

during later transport within the planet. Assuming no fractionation during transport from a deep interior reservoir to the MORB source region, $F_{36}^{\odot}$ would be ~1 if these gases originated from solar-wind implantation and retention in proto-Earth materials. Alternatively they could have been incorporated by adsorption of nebular gases on Earth's accretional core or into pre-planetary matter⁴. In this case, again without transport fractionation, one might expect $F_{36}^{\odot}$ to fall within the range of fractionation factors seen in laboratory adsorption experiments, from roughly 10 to 50 with a mean of ~25 (adsorption data are summarized and referenced in ref. 4). But as $F_{36}^{\odot}$ is unknown, it is used here as an adjustable parameter in fitting the mixing model to the measured trend in Fig. 1.

The parameter F_{36}^{a} , defined as $(^{36}\text{Ar}/^{22}\text{Ne})_{\text{sample}}/(^{36}\text{Ar}/^{22}\text{Ne})_{\text{a}}$, is taken to be 1. Patterson *et al.*¹⁷ argue that incorporation of an atmospheric contaminant probably results from equilibrium of basaltic melts with noble gases dissolved in sea water. The product of fractionations between air and seawater gases ($(^{36}\text{Ar}/^{22}\text{Ne})_{\text{seawater}}/(^{36}\text{Ar}/^{22}\text{Ne})_{\text{a}} = 4.0$; ref. 18) and seawater and melt gases ($(^{36}\text{Ar}/^{22}\text{Ne})_{\text{melt}}/(^{36}\text{Ar}/^{22}\text{Ne})_{\text{seawater}} = 0.25$; ref. 19) yields a net F_{36}^{a} fractionation factor close to unity. But this assumption, that the atmospheric elemental ratio represents the air-derived contaminant in ocean basalt samples, is unlikely to be always valid. If, for example, atmospheric noble gases were adsorbed on sample surfaces after collection, the process could plausibly fractionate Ne and Ar in favour of the heavier Ar, yielding $F_{36}^{\text{a}} > 1$.

Modelling results are plotted in Fig. 2a as mixing trajectories superimposed on the Fig. 1 data field. The three solid curves between atmosphere and the central and extreme values of the solar-wind $^{38}\text{Ar}/^{36}\text{Ar}$ range are all satisfactory fits to the EPR data trend. $F_{36}^{\odot}$ values labelling each curve are those producing the best statistical matches to the measurements. They are reasonably well defined—variations of ±30% produce observably worse fits—and fall within the range of ≈10–50 characteristic of adsorptive fractionation. The dashed curve for $F_{36}^{\odot} = 1$ clearly fails; there is no value of $(^{38}\text{Ar}/^{36}\text{Ar})_{\odot}$ remotely close to the solar-wind range that can be used in equation (2) to fit the data if the solar source reservoir for these MORBs is elementally unfractionated.

Figure 2b shows the Loihi measurements¹¹ overlaid on the Fig. 2a mixing curves. Roughly one-third of the data points are consistent with the trend defined by the EPR samples. The others lie above it,

with several falling along a $F_{36}^{\odot} = 5$ curve. This Loihi distribution could imply heterogeneous elemental fractionation in the solar source reservoir. A plausible alternative, however, is that some of these samples have acquired an air contaminant fractionated in favour of Ar, with $F_{36}^{\text{a}} > 1$. The $F_{36}^{\odot} = 5$, $F_{36}^{\text{a}} = 1$ mixing curve is identical to that generated by taking $F_{36}^{\odot} = 22.5$ (the median match to the EPR data) and $F_{36}^{\text{a}} = 4.5$, close to the value expected for occluded seawater noble gases¹⁸. A more extreme fractionation of $F_{36}^{\odot} \sim 20$, suggesting adsorption directly from air, would, as shown, account nicely for the four data points trending to the right from the "Atmosphere" endpoint (Fig. 2b), as well as for the one similar EPR datum in Fig. 2a.

$^{40}\text{Ar}/^{36}\text{Ar}$ ratios of ≤6,000 from the EPR and Loihi measurements are plotted against $^{38}\text{Ar}/^{36}\text{Ar}$ in Fig. 2c. Scatter on this correlation diagram is a signature of variable multi-component mixing, in this case of air with two additional radiogenic ^{40}Ar components ($^{40}\text{Ar}_{\text{rad}}$) plausibly associated with upper- and lower-mantle reservoirs. The EPR and Loihi data tell a notably consistent story for $^{40}\text{Ar}_{\text{rad}}$. Several points from both data sets fall close to the upper MORB line¹¹ and also along a reasonably well-defined correlation bounding the distribution on the left, assumed here to represent mixing of air with a MORB-free lower-mantle component. Most of the remaining points lie between these boundaries. Air-free ratios of radiogenic to solar Ar in the MORB and lower-mantle reservoirs are estimated by extrapolating these correlations¹¹, here to their intersections with the limits of the $(^{38}\text{Ar}/^{36}\text{Ar})_{\odot}$ range in Table 1, yielding $^{40}\text{Ar}_{\text{rad}}/^{36}\text{Ar}_{\odot}$ values of ~66,000–110,000 for the MORB source, and ~3,800–6,200 for the lower-mantle reservoir. These high MORB ratios are reasonable extensions of the Burnard *et al.*¹³ estimate of ≥40,000 from popping rock, given their suspicion that most of the ^{36}Ar in their study was adsorbed air contamination.

Several conclusions may be drawn from the observational and modelling evidence for isotopically solar Ne and Ar components in these ocean basalts, their elemental fractionation with respect to the solar Ne/Ar ratio, and their association with what is thought to be upwellings of plume material from deep in the mantle^{11,12}.

First, the tendency of non-radiogenic Ar, Kr and Xe in most mantle samples to display approximately atmospheric compositions is probably due to massive air-derived contamination by post-collection adsorption¹³, seawater–magma interaction¹⁷,

subduction²⁰, or combinations of all of these, which mask solar signatures.

Second, the failure of elementally unfractionated ($F_{36}^{\odot} \approx 1$) solar gases to provide even approximate modelling fits to the data trends appears to rule out direct implantation of solar-wind ions into terrestrial accretional matter as a likely mechanism for initially populating a deep, primordial mantle reservoir. The $F_{36}^{\odot}$ values that do generate matches over the uncertainty range in $(^{38}\text{Ar}/^{36}\text{Ar})_{\odot}$ bracket the average adsorptive fractionation factor of ~ 25 observed

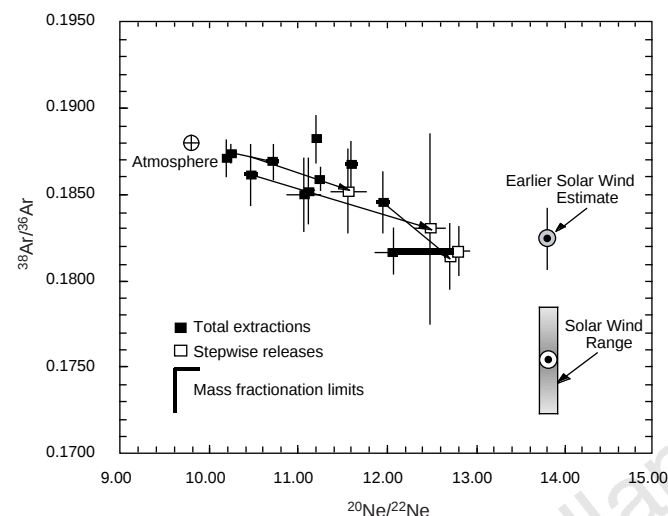


Figure 1 Correlation diagram of neon and argon isotope ratios measured in MORB glasses from the southern EPR¹². Data from Table 1, and from Benkert *et al.*¹⁶ for the earlier estimate of solar-wind composition. The plotted MORB data, taken directly from Tables 1 and 2 in ref. 12, include both total extraction compositions (filled symbols) and those in selected thermal release fractions (open symbols) which display extreme Ne and Ar isotope ratios. Three of the four release fraction compositions in Table 1 are from single temperature steps or summations of Ne and Ar in two consecutive steps. The exception is the integrated 1,600°C plus 2,000°C releases from the 'incompletely degassed' G296G* sample, which include only the Ne evolved at 1,600°C; Ne in the 2,000°C fraction is very uncertain in amount¹² and is isotopically consistent with a laboratory air contaminant. For this reason, and because he regards the Ar blank corrections in the 2,000°C step as very uncertain, S. Niedermann (personal communication) recommends its exclusion from the data base. The G296G* data, connected by a heavy horizontal bar in this figure, were therefore given no weight in fitting the Fig. 2a mixing curves to the measured isotopic distribution. The fits are in fact improved without them because of the aberrant position of the 'total extraction' point (filled square). It is necessary to consider the possibility of isotopic mass fractionation in partial extractions—that is, in thermal release fractions and 'incompletely degassed' samples: the trends of these data with respect to the total isotopic compositions of their parent samples are in the right direction. These fractions are connected to their parents (or, in the case of 'incompletely degassed' samples, to compositions in totally extracted splits of their parents¹²) by arrowed lines in this figure. Fractionations governed by Fick's Law diffusion from uniformly loaded mineral grains are restricted to $(m_1/m_2)^{1/4}$ —or less, depending on the fraction of gas evolved—with respect to total grain compositions. The orthogonal bars labelled 'mass fractionation limits' represent these maximum isotopic fractionations in the two coordinate directions. The observation that displacements of partial extractions from their parent compositions along the $^{20}\text{Ne}/^{22}\text{Ne}$ direction are considerably larger rules out fractionation as a major contributor to the data trend, at least for Ne. This conclusion is less firm for $^{38}\text{Ar}/^{36}\text{Ar}$, with its relatively much smaller variations and larger uncertainties. Nevertheless the fact that nine of the ten (excluding G296G*) total extraction points—which are not subject to fractionation effects—fall on the Fig. 2a mixing curves within $\leq 1\sigma$, and that the partial release data also lie closely along these same curves, strongly suggests a distribution trend due primarily to two-component mixing for both gases.

in laboratory experiments, pointing to adsorption of nebular gases as a likely incorporation mechanism⁴. These elevations of Ar/Ne relative to the solar ratio are of the same order as the factors of ~ 15 – 30 inferred for interior planetary reservoirs in theoretical models of atmospheric evolution on Earth and Mars^{4,5}, and of ~ 10 and ~ 35 estimated respectively by Burnard *et al.*¹³ and Moreira *et al.*²¹ from measurements on popping rock 2πD43. The high Kr/Ne and Xe/Ne ratios found by Moreira *et al.*²¹ are also qualitatively consistent with adsorptive fractionation of an elementally solar composition, but here the implied fractionation factors are substantially larger, particularly for Kr/Ne, than one would expect from laboratory experiments⁴. An alternative interpretation is suggested by the observation that these relative abundances are a rather good match to the air pattern, particularly if some preferential take-up of air Xe has occurred. So it may be that most of the Kr and Xe in popping rock, and some fraction of the Ar as well¹³, are atmospheric in origin—a possibility recognized by Moreira *et al.*²¹.

Third, appearance of solar Ar signatures in these sample suites has important implications for the origin and evolution of the terrestrial atmosphere. The atmosphere has historically been widely regarded as secondary—that is, degassed from the interior after complete dissipation of a primordial atmosphere (if any) that may have formed during planetary accretion, and subsequently unaltered except for outgassing of radiogenic species such as ^{40}Ar . This presumption, which requires atmospheric compositions for non-radiogenic noble gases in the interior reservoir, was called into question with the discovery of solar-like Ne in the upper mantle. One approach to the Ne problem invokes limited modification of the degassed atmosphere by an early episode of escape to space^{3,22} which depleted and fractionated Ne but not heavier species, while retaining the assumption of air-like compositions for Ar–Kr–Xe in the interior sources from which the atmosphere outgassed (see, for example, ref. 23). However the data discussed here point to degassing of Ar that was isotopically solar, not atmospheric, and so it too must have been fractionated by atmospheric escape. The data also argue against models in which the Earth acquired its interior Ar, Kr and Xe from accretion of carbonaceous meteorites carrying the so-called 'planetary' noble gas pattern²⁴; average $^{38}\text{Ar}/^{36}\text{Ar}$ in the CI-meteorites is indistinguishable from the air ratio⁴ and therefore far from the solar-like composition in the EPR and Loihi source reservoirs.

Detection of solar Ar in the mantle supports modelling arguments for an initial terrestrial inventory that was solar for all the noble gases^{3–5,13}. In this view, atmospheric noble gases have been isotopically altered from their source composition, and thus are not secondary in the usual sense. The fact that non-radiogenic atmospheric Ne, Ar, Kr and Xe are all isotopically heavier than their solar counterparts is an important clue to the nature of the alteration process, and is central to evolutionary models in which the atmosphere was driven from solar progenitors to its present compositional state by energetic escape of a primordial atmosphere, followed by mixing of its fractionated residue with gases evolved from the interior^{4,5}.

If the models reported here are approximately correct, isolation of a solar mantle component in Xe, where atmospheric and solar light-isotope ratios are very different²⁵, is only a matter of sampling luck and analytic precision. The associated Kr signature will probably be harder to recognize because of the comparatively small isotopic difference between the air and solar compositions^{6,25}. □

Received 19 March; accepted 13 May 1998.

- Honda, M., McDougall, I., Patterson, D. B., Doulgeris, A. & Clague, D. A. A solar component in the Earth: Neon isotope anomalies in Loihi and Kilauea basalts. *Nature* **349**, 149–151 (1991).
- Honda, M., McDougall, I., Patterson, D. B., Doulgeris, A. & Clague, D. A. Noble gases in submarine pillow basalt glasses from Loihi and Kilauea, Hawaii: A solar component in the Earth. *Geochim. Cosmochim. Acta* **57**, 859–874 (1993).
- Farley, K. A. & Poreda, R. J. Mantle neon and atmospheric contamination. *Earth Planet. Sci. Lett.* **114**, 325–339 (1993).
- Pepin, R. O. On the origin and early evolution of terrestrial planet atmospheres and meteoritic volatiles. *Icarus* **92**, 2–79 (1991).

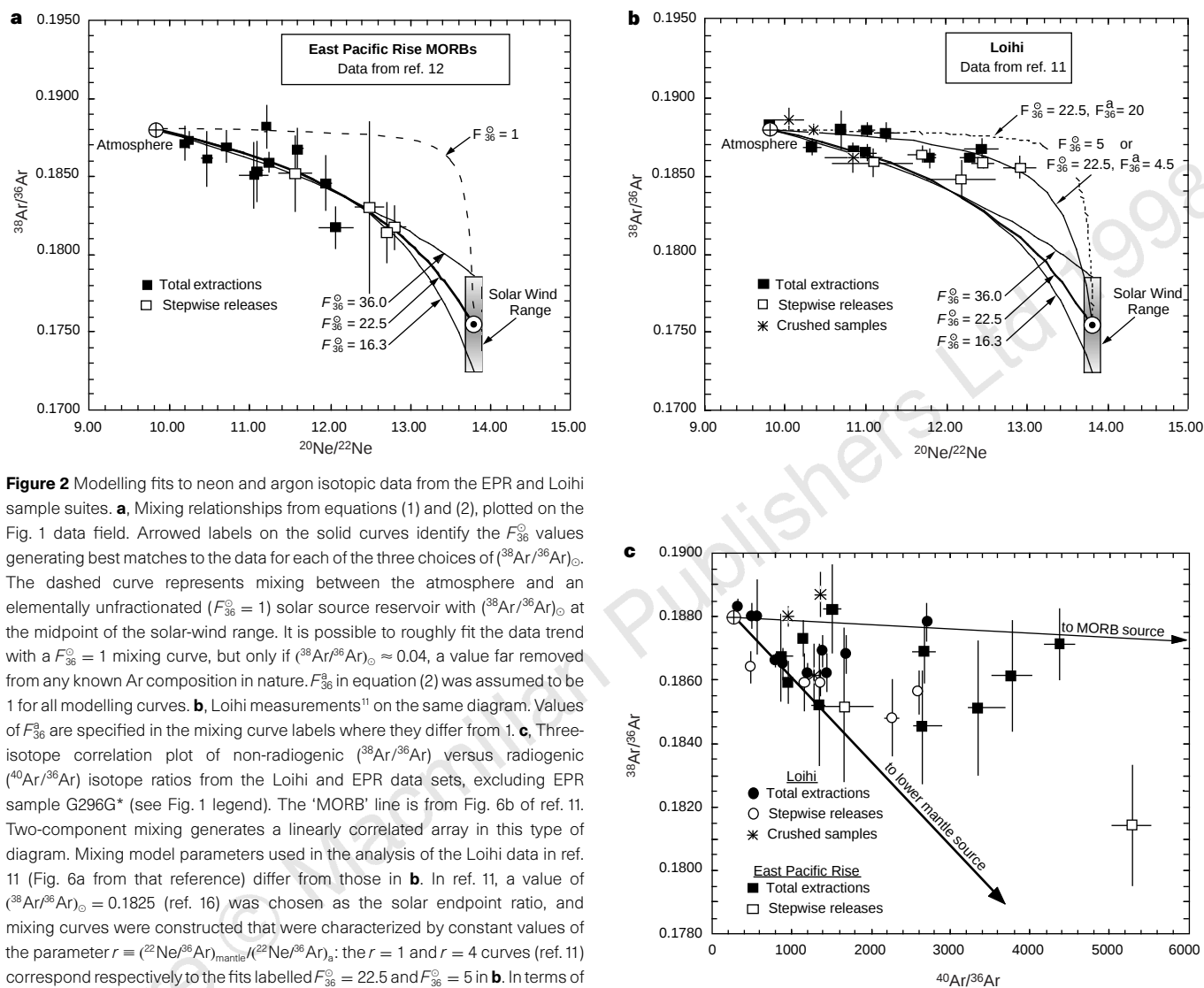


Figure 2 Modelling fits to neon and argon isotopic data from the EPR and Loihi sample suites. **a**, Mixing relationships from equations (1) and (2), plotted on the Fig. 1 data field. Arrowed labels on the solid curves identify the F_{36}^o values generating best matches to the data for each of the three choices of $(^{38}\text{Ar}/^{36}\text{Ar})_o$. The dashed curve represents mixing between the atmosphere and an elementally unfractionated ($F_{36}^o = 1$) solar source reservoir with $(^{38}\text{Ar}/^{36}\text{Ar})_o$ at the midpoint of the solar-wind range. It is possible to roughly fit the data trend with a $F_{36}^o = 1$ mixing curve, but only if $(^{38}\text{Ar}/^{36}\text{Ar})_o \approx 0.04$, a value far removed from any known Ar composition in nature. F_{36}^o in equation (2) was assumed to be 1 for all modelling curves. **b**, Loihi measurements¹¹ on the same diagram. Values of F_{36}^o are specified in the mixing curve labels where they differ from 1. **c**, Three-isotope correlation plot of non-radiogenic ($^{38}\text{Ar}/^{36}\text{Ar}$) versus radiogenic ($^{40}\text{Ar}/^{36}\text{Ar}$) isotope ratios from the Loihi and EPR data sets, excluding EPR sample G296G* (see Fig. 1 legend). The 'MORB' line is from Fig. 6b of ref. 11. Two-component mixing generates a linearly correlated array in this type of diagram. Mixing model parameters used in the analysis of the Loihi data in ref. 11 (Fig. 6a from that reference) differ from those in **b**. In ref. 11, a value of $(^{38}\text{Ar}/^{36}\text{Ar})_o = 0.1825$ (ref. 16) was chosen as the solar endpoint ratio, and mixing curves were constructed that were characterized by constant values of the parameter $r = (^{22}\text{Ne}/^{36}\text{Ar})_{\text{mantle}} / (^{22}\text{Ne}/^{36}\text{Ar})_a$; the $r = 1$ and $r = 4$ curves (ref. 11) correspond respectively to the fits labelled $F_{36}^o = 22.5$ and $F_{36}^o = 5$ in **b**. In terms of the parameters used here, r is expressed as $[(^{36}\text{Ar}/^{22}\text{Ne})_o F_{36}^o]^{-1} (^{36}\text{Ar}/^{22}\text{Ne})_a$, and is equal to 2.8 and 12.7, compared to values of $r = 1$ and $r = 4$ in ref. 11. These differences are due entirely to the lower $(^{38}\text{Ar}/^{36}\text{Ar})_o$ range adopted here for the solar wind (see Fig. 1 and text). They imply atmospheric losses of Ne, relative to Ar, of ~65% and 90%, compared with 0% and 75% for the $r = 1$ and $r = 4$ cases, respectively, in ref. 11).

- Pepin, R. O. Evolution of Earth's noble gases: Consequences of assuming hydrodynamic loss driven by giant impact. *Icarus* **126**, 148–156 (1997).
- Pepin, R. O. Are isotopically solar components present in heavy noble gases from terrestrial mantle samples? Perhaps ... *Lunar Planet. Sci. XXVI*, 1109–1110 (Lunar & Planetary Institute, Houston, 1995).
- Ozima, M. & Zashu, S. Noble gas state of the ancient mantle as deduced from noble gases in coated diamonds. *Earth Planet. Sci. Lett.* **105**, 13–27 (1991).
- Ozima, M. Noble gas state in the mantle. *Rev. Geophys.* **32**, (4) 405–426 (1994).
- Hiyagon, H., Ozima, M., Marty, B., Zashu, S. & Sakai, H. Noble gases in submarine glasses from mid-oceanic ridges and Loihi seamount: Constraints on the early history of the Earth. *Geochim. Cosmochim. Acta* **56**, 1301–1316 (1992).
- Poreda, R. J. & Farley, K. A. Rare gases in Samoan xenoliths. *Earth Planet. Sci. Lett.* **113**, 129–144 (1992).
- Valbracht, P. J., Staudacher, T., Malahoff, A. & Allègre, C. J. Noble gas systematics of deep rift zone glasses from Loihi Seamount, Hawaii. *Earth Planet. Sci. Lett.* **150**, 399–411 (1997).
- Niedermann, S., Bach, W. & Erzinger, J. Noble gas evidence for a lower mantle component in MORBs from the southern East Pacific Rise: Decoupling of helium and neon isotope systematics. *Geochim. Cosmochim. Acta* **61**, 2697–2715 (1997).
- Burnard, P., Graham, D. & Turner, G. Vesicle-specific noble gas analyses of "popping rock": Implications for primordial noble gases in Earth. *Science* **276**, 568–571 (1997).
- Becker, R. H., Schlutter, D. J., Rider, P. E. & Pepin, R. O. An acid-etch study of the Kapoeta achondrite: Implications for the argon-36/argon-38 ratio in the solar wind. *Meteorit. Planet. Sci.* **33**, 109–113 (1998).
- Pepin, R. O. & Schlutter, D. J. Measurement of neon and argon isotopic compositions in single lunar regolith grains. *Meteorit. Planet. Sci.* **32**, A104–A105 (1997).
- Benkert, J.-P., Baur, H., Signer, P. & Wieler, R. He, Ne, and Ar from the solar wind and solar energetic particles in lunar ilmenites and pyroxenes. *J. Geophys. Res.* **98**, 13147–13162 (1993).
- Patterson, D. B., Honda, M. & McDougall, I. Atmospheric contamination: A possible source for heavy

- noble gases in basalts from Loihi seamount, Hawaii. *Geophys. Res. Lett.* **17**, 705–708 (1990).
- Allègre, C. J., Staudacher, T. & Sarda, P. Rare gas systematics: Formation of the atmosphere, evolution and structure of the Earth's mantle. *Earth Planet. Sci. Lett.* **81**, 127–150 (1986).
- Lux, G. The behavior of noble gases in silicate liquids: Solution, diffusion, bubbles and surface effects, with applications to natural samples. *Geochim. Cosmochim. Acta* **51**, 1549–1560 (1987).
- Porcelli, D. & Wasserburg, G. J. Mass transfer of xenon through a steady-state upper mantle. *Geochim. Cosmochim. Acta* **59**, 1991–2008 (1995).
- Moreira, M., Kunz, J. & Allègre, C. J. Rare gas systematics in popping rock: Isotopic and elemental compositions in the upper mantle. *Science* **279**, 1178–1181 (1998).
- Zahnle, K., Kasting, J. & Pollack, J. Mass fractionation of noble gases in diffusion-limited hydrodynamic hydrogen escape. *Icarus* **84**, 502–527 (1990).
- Ozima, M. Primordial terrestrial noble gases and Earth evolution. *Eos* **78**, F810 (1997).
- Harper, C. L. Jr & Jacobsen, S. B. Noble gases and Earth's accretion. *Science* **273**, 1814–1818 (1996).
- Pepin, R. O., Becker, R. H. & Rider, P. E. Xenon and krypton isotopes in extraterrestrial regolith soils and in the solar wind. *Geochim. Cosmochim. Acta* **59**, 4997–5022 (1995).
- Eberhardt, P., Eugster, O. & Marti, K. A redetermination of the isotopic composition of atmospheric neon. *Z. Naturforsch.* **20a**, 623–624 (1965).
- Nier, A. O. A redetermination of the relative abundances of the isotopes of carbon, nitrogen, oxygen, argon and potassium. *Phys. Rev.* **77**, 789–793 (1950).
- Ozima, M. & Podosek, F. A. *Noble Gas Geochemistry* (Cambridge Univ. Press, 1983).
- Murcr, C. A., Baur, H., Signer, P. & Wieler, R. Helium, neon, and argon abundances in the solar wind: In vacuo etching of meteoritic iron-nickel. *Geochim. Cosmochim. Acta* **61**, 1303–1314 (1997).

Acknowledgements. I thank M. Honda for suggestions, and S. Niedermann for additional information about his EPR data. This work was supported by the Program of NASA.

Correspondence and requests for materials should be addressed to the author (e-mail: pepin001@tc.umn.edu).

Evolution of an active sea-floor massive sulphide deposit

C.-F. You*† & M. J. Bickle*

* Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

† Department of Earth Sciences, National Cheng-Kung University, Tainan, Taiwan

Hydrothermal circulation at oceanic spreading ridges causes sea water to penetrate to depths of 2 to 3 km in the oceanic crust where it is heated to ~400 °C before venting at spectacular 'black smokers'. These hydrothermal systems exert a strong influence on ocean chemistry¹, yet their structure, longevity and magnitude remain largely unresolved². The active Transatlantic Geotraverse (TAG) deposit, at 26° N on the Mid-Atlantic Ridge, is one of the largest, oldest and most intensively studied of the massive sulphide mounds that accumulate beneath black-smoker fields. Here we report ages of sulphides and anhydrites from the recently drilled³ TAG substrate structures—determined from ²³⁴U–²³⁰Th systematics analysed by thermal ionization mass spectrometry. The new precise ages combined with existing data^{4,5} show that the oldest material (11,000 to 37,000 years old) forms a layer across the centre of the deposit with younger material (2,300–7,800 years old) both above and below. This stratigraphy confirms that much of the sulphide and anhydrite are precipitated within the mound by mixing of entrained sea water with hydrothermal fluid⁶. The age distribution is consistent with episodic activity of the hydrothermal system recurring at intervals of up to 2,000 years.

The uncertainty over the magnitude and temperature distribution of the flux from oceanic hydrothermal systems arises because the structure of the hydrothermal systems is not well known and because surface observations have only been made over a period of less than 20 years. The water may either be heated in a boundary layer within tens of metres of an underlying convecting magma chamber⁷ or in a dynamic situation in a 'cracking front' which rapidly penetrates a recently solidified magma body⁸. The magnitude of the high-temperature flux from active black-smoker com-

plexes requires that high-temperature venting is intermittent^{9,10}. The high-temperature episodes might be expected to last hundreds to thousands of years above a ~100-m-thick magma chamber to as little as weeks above individual ~1-m-wide dykes¹¹. The complexity of the processes within sea-floor sulphide deposits has been confirmed by the results of the Leg 158 of the Ocean Drilling Program which show that the sulphide deposit at TAG is up to ~150 m thick and has a complex internal stratigraphy (Fig. 1)³. The presence of anhydrite, predicted on the basis of effluent chemistry⁶, attests to the extent of interactions between sea water and hydrothermal fluid within the mound. Previous ²³⁴U–²³⁰Th age determinations on material dredged from the mound surface and the new drill core show that the mound is ~50,000 years old and that some samples deep in the mound are relatively young^{4,5}. These ages, determined by α -spectrometry, typically have 2 σ errors between 10% and 25% and sample-size requirements necessitated analysis of bulk samples of potentially polygenetic anhydrite–pyrite breccias.

U–Th isotopic data and ages on 24 pyrite or anhydrite separates from Leg 158 drill core are listed in Table 1 and shown in Fig. 1, combined with data from Lalou *et al.*^{4,5}. The striking results are as follows: (1) that the distribution of ages correlates with the stratigraphy proposed by Humphris *et al.*³; (2) the oldest unit is the pyrite-silica breccia layer at mid-height in the mound; (3) the mound comprises material older than 2,200 years except on its surface (all of the samples we processed are included in Table 1); and (4) individual layers within the mound exhibit a significant spread of ages.

The correlation of the ages with the stratigraphy is a prime indication of the gross integrity of the ages, despite ²³⁴U–²³⁰Th sulphide ages being sensitive to a number of perturbations. ²³⁴U–²³⁰Th dating of young sulphides is based on ingrowth of ²³⁰Th (half-life ~75,000 years) from decay of ²³⁴U in samples with negligible initial Th (ref. 12). The U–Th systematics may be perturbed by U being subsequently fixed at oxidation–reduction boundaries on pyrites¹³ or lost from this mineral. In addition, newly crystallizing minerals might have incorporated reworked U or Th from older deposits within the mound. Some ²³⁴U–²³⁰Th ages from TAG have been checked against ²²⁶Ra/²³⁴U ages⁵ and a ²³¹Pa/²³⁵U age⁴, and agree within the analytical errors of 10–20%. Assessment of the integrity of the ages at higher precisions is more problematic. Concentrations of ²³²Th, with one exception, are less than 2 ng g⁻¹ and incorporation of Th from basalt or sediment would have a negligible effect on ²³⁰Th/²³⁴U ratios. Initial ²³⁴U/²³⁸U ratios of the

Table 1 U and Th isotope analyses

Sample*	Min.†	U (ng g ⁻¹)	1 σ (%)	Th (ng g ⁻¹)	1 σ (%)	²³⁴ U _{act.} ²³⁸ U	1 σ (%)	²³⁴ U _{act.} ± ²³⁸ U _i	²³⁰ Th _{atom} ²³² Th	1 σ (%)	²³⁰ Th _{act.} ²³⁸ U	1 σ (%)	Ag (yr)	1 σ
C-12N-1-40-42a	An	227	0.10	0.879	0.55	1.281	6.11	1.290	0.000516	5.75	0.1184	6.02	10,507	920
C-12N-1-40-42b	Py	1,997	0.23	0.276	1.63	1.144	1.13	1.146	0.004418	5.10	0.0372	3.96	3,592	150
C-14N-1-9-10a	An	2,893	0.12	1.784	0.48	1.130	1.32	1.134	0.002691	1.69	0.1031	1.46	10,385	214
C-14N-1-9-10b	Py	1,069	0.14	0.076	0.26	1.126	0.40	1.128	0.018037	0.60	0.0774	0.40	7,736	45
C-14N-1-68-70a	An	776	0.14	1.508	0.80	1.132	4.82	1.134	0.000332	5.19	0.0406	5.77	3,972	299
C-14N-1-68-70b	Py	4,716	0.11	0.102	0.58	1.135	0.25	1.137	0.028673	3.30	0.0378	3.54	3,679	132
E-5R-1-34-36	Py	1,751	0.19	0.496	0.41	1.160	0.50	1.167	0.008584	0.64	0.1477	0.74	14,767	142
E-8R-1-16-20	Py	2,758	0.08	1.650	0.33	1.135	0.19	1.137	0.001485	3.85	0.0540	3.05	5,294	165
E-15R-1-12-14	Py	7,368	0.39	0.575	1.53	1.137	0.42	1.139	0.011688	1.82	0.0548	1.03	5,369	61
G-3N-1-44-47a	Py	6,258	0.60	0.539	0.27	1.139	0.45	1.140	0.006364	0.35	0.0334	0.67	3,231	27
G-3N-1-44-47b	Py	6,280	1.11	0.380	0.27	1.151	1.55	1.152	0.007772	0.76	0.0287	1.25	2,746	55
G-3N-1-44-47c	Py	8,780	0.16	1.881	0.34	1.150	0.11	1.152	0.002257	0.46	0.0294	0.36	2,817	10
G-3N-1-44-47d	An	414	0.25	4.947	0.28	1.136	2.03	1.137	0.000031	0.80	0.0288	8.61	2,797	251
G-3N-1-44-47e	Py	6,450	0.47	0.594	0.73	1.146	0.40	1.147	0.004803	0.16	0.0274	2.18	2,633	59
G-3N-1-44-47	Py	6,310	0.62	0.471	0.68	1.164	0.71	1.166	0.007936	0.82	0.0326	1.63	3,086	56
H-5N-1-44-51	Py	2,022	0.13	0.693	0.25	1.135	0.44	1.139	0.005126	0.69	0.1073	0.57	10,789	83
H-5N-1-0-3a	Py	303	0.18	0.241	0.33	1.149	0.82	1.167	0.005398	0.46	0.3423	2.60	38,150	1,246
H-5N-1-0-3b	Py	313	0.28	0.352	1.55	1.195	1.03	1.216	0.004939	2.89	0.3386	2.22	35,895	1,040
H-8N-1-90-92	Py	7,765	1.07	1.215	0.78	1.137	0.80	1.140	0.007183	0.44	0.0686	1.23	6,761	102
I-1N-1-6-8	Py	2,892	1.64	0.342	1.26	1.272	2.44	1.278	0.011805	3.11	0.0843	2.80	7,439	283
K-1X-1-30-33	Py	8,934	0.29	0.413	2.58	1.141	0.58	1.144	0.020788	2.67	0.0576	1.95	5,629	118
M-2R-1-15-17	Py	4,155	0.21	0.331	0.42	1.140	0.46	1.143	0.015514	0.44	0.0745	0.40	7,344	
M-3R-1-60-64	Py	5,360	0.10	1.151	0.10	1.140	0.49	1.144	0.008935	0.11	0.1166	0.16	11,722	64
M-8R-1-8-10	Py	22,535	0.51	1.197	0.5	1.154	0.23	1.156	0.013547	0.42	0.0433	0.76	4,150	34

In column headings; 'act.' indicates activity ratio, 'atom' indicates atom ratio; 1 σ indicates error of measurement in preceding column.

* Sample numbers comprise: site 957 Hole (letter), type, section no, depth interval in tray in cm (ref. 16), our sub-sample letter.

† Mineralogy: An, anhydrite; Py, pyrite.

‡ ²³⁴U_{act.}/²³⁸U_i is initial ²³⁴U/²³⁸U activity ratio.

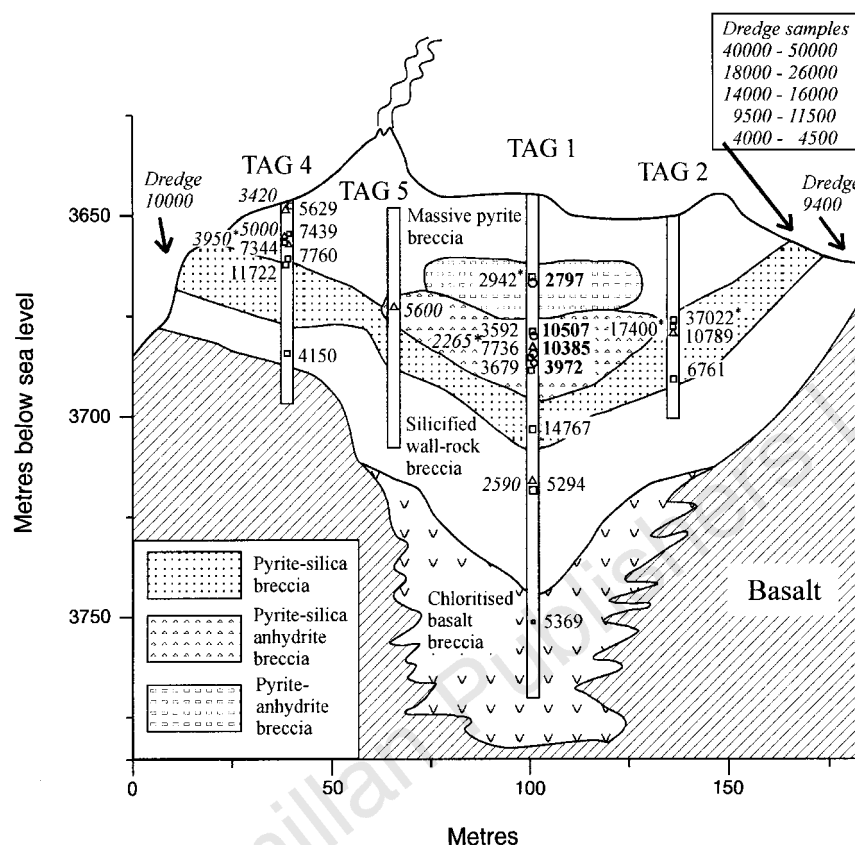


Figure 1 Stratigraphy of the TAG mound modified after Humphris *et al.*³. The section is drawn northwest-southeast, approximately through centre of mound with topography from ref. 17. Drill sites (for example, TAG1) have been projected onto the section, and multiple drill holes have been combined into one schematic hole. Ages from this study are shown in normal type (locations shown by open squares), with ages on anhydrite mineral fractions in bold (open circles). Ages

from Lalou *et al.*⁵ are shown in italic (open triangles). Ages on dredge samples are taken from Lalou *et al.*⁴. Asterisk indicated an average of >1 age. The central pyrite-silica breccia, characterized by ages >11,000 years, has been extended to the surface because dredge samples of this age range have been recovered on the surface on both the northwestern and southeastern perimeter of the mound.

sulphides are, with two exceptions, within $\pm 1\%$ of the ratio in sea water, which is unlikely to have varied significantly over the past 40,000 years. Elevation of initial $^{234}\text{U}/^{238}\text{U}$ ratios is characteristic of carbonate samples subject to post-formation disturbances¹⁴. Six sub-samples of G-3N-1-44-47 scatter about an isochron in excess of analytical error (Fig. 2). This gives an initial $^{230}\text{Th}/^{232}\text{Th}$ ratio within error of the ratio of crustal materials ($\sim 10^{-5}$), confirming that the sample did not incorporate significant ^{230}Th from older mound material.

The spread in ages of adjacent samples (Figs 1 and 2) is interpreted to result from the complex seawater-hydrothermal fluid interactions involved in precipitation of sulphides and anhydrite within the mound. Laser inductively coupled plasma mass spectrometry analyses show that the TAG pyrites from the samples exhibit 100-fold variations of U content on a sub-millimetre scale¹⁵. Given that the U is entirely derived from sea water (initial $^{234}\text{U}/^{238}\text{U}$ activity ratios, $1.149 \pm 8(2\sigma)$; Table 1), it is probable that pyrite precipitation results from mixing between sea water and hydrothermal fluid, with U strongly partitioning into the fluid phase. If so, the 1–3-g samples analysed in this work comprise many individual growth zones, and date the average age over which the bulk sample was precipitated. The implication is that the mixing between sea water and hydrothermal fluid took place in a permeable sulphide matrix in which the permeability survived or was recreated by dissolution of anhydrite over hundreds of years. The four anhydrite ages indicate that some anhydrite does survive for long periods in the mound, despite the inference from mass-balance calculations that most anhydrite is rapidly redissolved⁶. Ages of two of the anhydrites are within error of those of sulphide from the same sample (C-14N-1-68-70 and G-3N-1-44-47), but two shallower

samples (from C12N-1-40-42 and C14N-1-9-10) contain $\sim 10,500$ -year-old anhydrite coexisting with 3,592- and 7736-year-old pyrites.

Previous compilations of sulphide ages have been interpreted in terms of episodic high-temperature activity^{4,5} consistent with the heat-flux constraints¹⁰. However, the previous sampling was

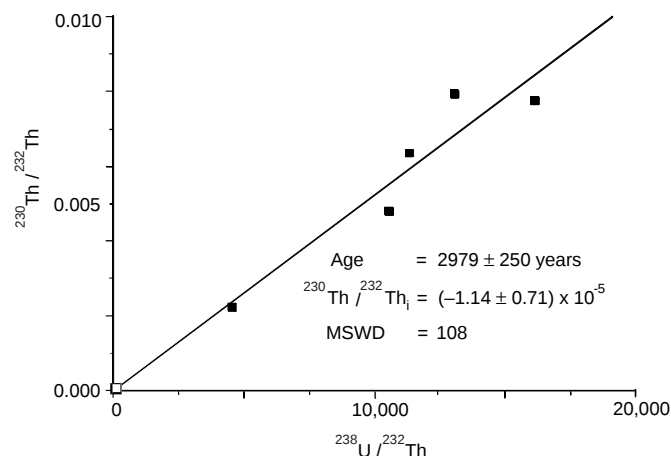


Figure 2 $^{230}\text{Th}/^{232}\text{Th}$ - $^{238}\text{U}/^{232}\text{Th}$ isochron diagram of five pyrite samples (filled squares) and one anhydrite sample (open square) from 957G-3N-1-44-47. Errors shown as 2σ . Samples scatter around an errorchron (mean squared weighted deviate = 108) with intercept within $\pm 10^{-5}$ of zero, indicating that Th incorporated at crystallization had a $^{230}\text{Th}/^{232}\text{Th}$ ratio typical of crustal materials. Regression after York¹⁸.

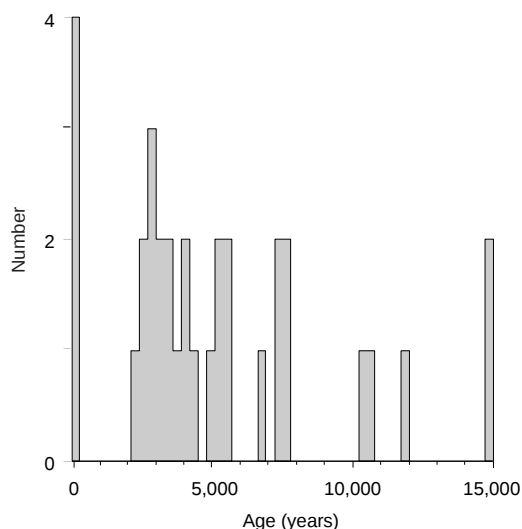


Figure 3 Frequency-age diagram for TAG sulphide and anhydrite samples with 2σ errors <600 years. Data are from this study and Lalou *et al.*⁵. Bin size, 300 years. We note that samples in the zero-age bin are over-represented because of their accessibility on the surface.

restricted to the surface, and contained relatively few ages; these ages had errors comparable with the potential $\sim 1,000$ -year episodicity of the system. A compilation of the ages from this study and those samples reported by Lalou *et al.*⁵ with 2σ errors less than 600 years indicates that activity at TAG may have been episodic at intervals of $<2,000$ years (Fig. 3). Interpretation of age-frequency diagrams with limited data is problematic. Monte-Carlo simulation indicates that the probability of a set of ages spread randomly over 12,000 years grouping to the extent observed (≤ 5 groups with ≥ 2 members) is 12%. If the samples had constant probability of survival, such that the sample abundance decreased exponentially with age with an exponential constant of $1/12,000$ years, the chance of the observed grouping occurring, despite the ages being random, increases to 20%. Significantly more data are needed to determine the timing and duration of the high-temperature events. The period of inactivity between $\sim 2,300$ years (samples dated at $2,633 \pm 118$ and $2,265 \pm 268$ years) and the recent activity (which is younger than ~ 50 years) is statistically well established.

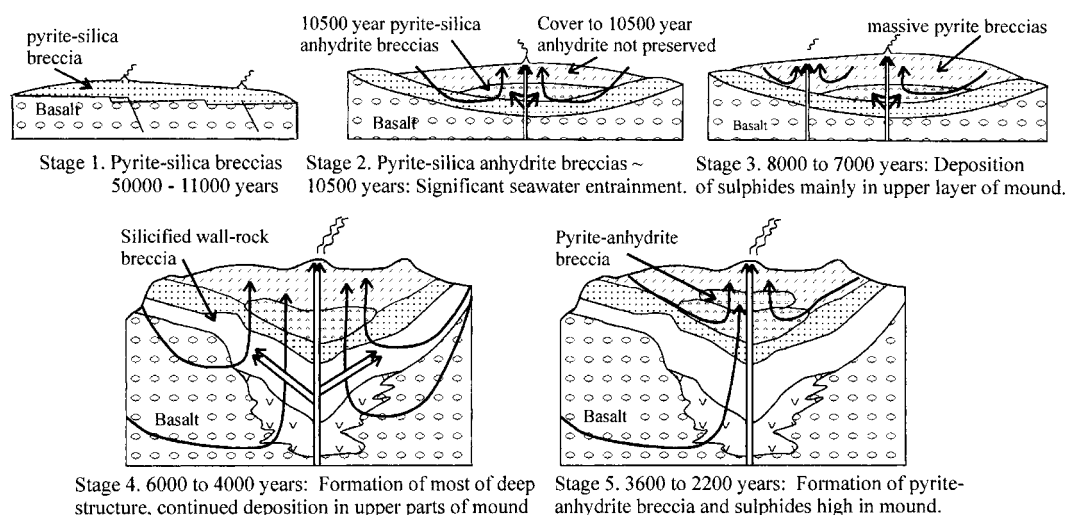


Figure 4 Schematic stages in the evolution of the TAG hydrothermal mound. We note that relatively few samples are available from the protracted first stage, and that significant wasting of the mound may have taken place at several intervals. Similarly, stage 2 anhydrite must have been deposited under a significant cover,

The distribution of ages within the mound indicates a protracted and complex evolution (Fig. 4). The first stage resulted in the pyrite-silica breccia layer with ages older than 11,000 yr, but the record of this surprisingly protracted stage is fragmentary. The subsequent stages (2 to 5) variably deposited pyrite and anhydrite both below and above the old pyrite-silica layer. Most, but not all, of the anhydrite was subsequently leached by circulation of sea water. The final (sixth) stage is the current activity which is younger than ~ 50 years (ref. 4) and only sampled on the surface.

The U-Th sulphide and anhydrite ages from the TAG drill core confirm the longevity of the mound, the stratigraphy erected on the basis of drill-core petrography and the complex evolution of the fluid flow geometry within the mound. Further work is necessary to refine the chronological evolution to the point at which individual high-temperature episodes can be confidently resolved and related to the structure of the reaction zone where fluid is heated in the crust.

Methods

U and Th concentrations and isotopic compositions were analysed by thermal ionization mass spectrometry on the T40 machine at Cambridge University. U and Th blanks were <30 pg and <20 pg, respectively. Th and U isotope ratios and concentrations were analysed on samples spiked with a mixed ^{233}U , ^{236}U and ^{229}Th spike. U was analysed either by ion counting, or by a combination of ion-counting and fixed Faraday collectors in a peak switching mode to calibrate Daly-Faraday gain. U analyses were fractionation corrected using both $^{235}\text{U}/^{238}\text{U}$ and $^{233}\text{U}/^{236}\text{U}$ ratios. 27 analyses of SRM U-72/05 gave $^{235}\text{U}/^{238}\text{U} = 0.99296 \pm 18(1\sigma)$. Th was analysed by ion-counting. Abundance sensitivity (without a filter) was better than 1 p.p.m. at 2 mass units, and tailing from ^{232}Th onto ^{230}Th was insignificant. U analyses, and the repeatability of Th analyses, indicates that fractionation of Th isotopic ratios was $<0.25\%$ per mass unit. Pyrite was separated from anhydrite and quartz by hand-picking and then leaching with 15 M HNO_3 . Anhydrite was leached in very dilute HCl, and quartz was discarded.

Received 22 December 1997; accepted 11 May 1998.

- Edmond, J. M. *et al.* Ridge crest hydrothermal activity and the balances of the major and minor elements in the ocean: The Galapagos data. *Earth Planet. Sci. Lett.* **46**, 1–18 (1979).
- Elderfield, H. & Elderfield, H. & Schultz, A. Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Annu. Rev. Earth Planet. Sci.* **24**, 191–224 (1996).
- Humphris, S. E. *et al.* The internal structure of an active sea-floor massive sulphide deposit. *Nature* **377**, 713–716 (1995).
- Lalou, C. *et al.* New age data for Mid-Atlantic ridge hydrothermal sites: TAG and Snake Pit chronology revisited. *J. Geophys. Res.* **98**, 9705–9713 (1993).
- Lalou, C., Reyss, J. L. & Brichet, E. Age of sub-bottom sulphide samples at the TAG active mound. *Proc. ODP Sci. Res.* **158**, 111–118 (1998).

none of which has been sampled. Present activity (<50 years) has relatively little effect on the mound, although young anhydrite and pyrite must be being deposited at depth; these deposits are not yet sampled.

6. James, R. H. & Elderfield, H. Chemistry of ore-forming fluids and mineral formation rates in an active hydrothermal sulfide deposit on the mid-Atlantic Ridge. *Geology* **24**, 1147–1150 (1996).
7. Cann, J. R., Strens, M. R. & Rice, A. A simple magma-driven thermal balance model for the formation of volcanogenic massive sulphides. *Earth Planet. Sci. Lett.* **76**, 123–134 (1985).
8. Wilcock, S. D. & Delany, J. R. Mid-ocean ridge sulfide deposits: evidence for heat extraction from magma chambers or cracking fronts? *Earth Planet. Sci. Lett.* **145**, 49–64 (1996).
9. Converse, D. R., Holland, H. D. & Edmond, J. M. Flow rates in the axial hot springs of the East Pacific Rise (21°N): implications for the heat budget and the formation of massive sulfide deposits. *Earth Planet. Sci. Lett.* **69**, 159–175 (1984).
10. Schultz, A., Delany, J. R. & McDuff, R. E. On the partitioning of heat flux between diffuse and point source venting. *J. Geophys. Res.* **97**, 12299–12314 (1992).
11. Lister, C. R. B. Heat transfer between magmas and hydrothermal systems, or, six lemmas in search of a theorem. *Geophys. J. Int.* **120**, 45–59 (1995).
12. Edwards, R. L., Chen, J. H. & Wasserburg, G. J. ^{238}U – ^{234}U – ^{230}Th – ^{232}Th systematics and the precise measurement of time over the past 500,000 years. *Earth Planet. Sci. Lett.* **81**, 175–192 (1986).
13. Mills, R., Thompson, J., Elderfield, H., Hinton, R. W. & Hyslop, E. Uranium enrichment in metalliferous sediments from the Mid-Atlantic Ridge. *Earth Planet. Sci. Lett.* **124**, 35–47 (1994).
14. Henderson, G. M., Cohen, A. S. & O'Nions, R. K. $^{234}\text{U}/^{238}\text{U}$ ratios and ^{230}Th ages for Hateruna Atoll corals: implications for coral diagenesis and seawater $^{234}\text{U}/^{238}\text{U}$ ratios. *Earth Planet. Sci. Lett.* **115**, 65–73 (1993).
15. You, C.-F., Bickle, M. J., Mills, R., Nesbitt, R. & Butler, I. Age and uranium mobility in hydrothermal sulphides. *Terra Nova* (Abstr. Suppl. 9) 555 (1997).
16. Humphris, S. E. *et al.* Explanatory notes. *Proc. ODP Init. Rep.* **158**, 37–53 (1996).
17. Humphris, S. E. *et al.* Introduction and principal results. *Proc. ODP Init. Rep.* **158**, 5–14 (1996).
18. York, D. Least squares fitting of a straight line with correlated errors. *Earth Planet. Sci. Lett.* **5**, 320–324 (1969).

Acknowledgements. We thank ODP and the staff at the Bremen core depository for collecting and curating samples; R. Mills, R. Nesbitt and I. Butler for discussions; and C. Lalou for making available a preprint of her work. This work was funded by NERC and the Newton Trust.

Correspondence and requests for materials should be addressed to M.J.B. (e-mail: mb72@esc.cam.ac.uk).

The gain of three mitochondrial introns identifies liverworts as the earliest land plants

Yin-Long Qiu*, Yangrae Cho, J. Colin Cox* & Jeffrey D. Palmer

Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

The first evidence for the emergence of land plants (embryophytes) consists of mid-Ordovician spore tetrads (~476 Myr old)^{1,2}. The identity of the early plants that produced these spores is unclear; they are sometimes claimed to be liverworts^{3,4}, but there are no associated megafossils, and similar spores can be produced by a diversity of plants². Indeed, the earliest unequivocal megafossils of land plants consist of early vascular plants and various plants of uncertain affinity¹. Different phylogenetic analyses have identified liverworts, hornworts and bryophytes as each being the first lineage of land plants^{1,2,5–13}; the consensus of these conflicting topologies yields an unresolved polychotomy at the base of land plants. Here we survey 352 diverse land plants and find that three mitochondrial group II introns are present, with occasional losses, in mosses, hornworts and all major lineages of vascular plants, but are entirely absent from liverworts, green algae and all other eukaryotes. These results indicate that liverworts are the earliest land plants, with the three introns having been acquired in a common ancestor of all other land plants, and have important implications concerning the early stages of plant evolution.

The three mitochondrial introns examined here, two from the *cox2* gene (introns *cox2.i3* and *cox2.i4*) and one from the *nad1* gene (*nad1.i4*), are known to be widespread in angiosperms but absent from the liverwort *Marchantia*^{14,15}. Because nucleotide substitutions in plant mitochondrial genomes occur generally at a low rate^{16,17}, probes internal to these three introns from angiosperms hybridized to nearly the full range of 352 land-plant DNAs examined. All major

groups of extant land plants were examined and, with the exception of liverworts, DNA from some members of all groups hybridized to at least two of the introns (Fig. 1). None of the 11 liverworts examined, which represent four of the six orders of liverworts, hybridized to any of the three introns (Figs 1 and 2). Because intron

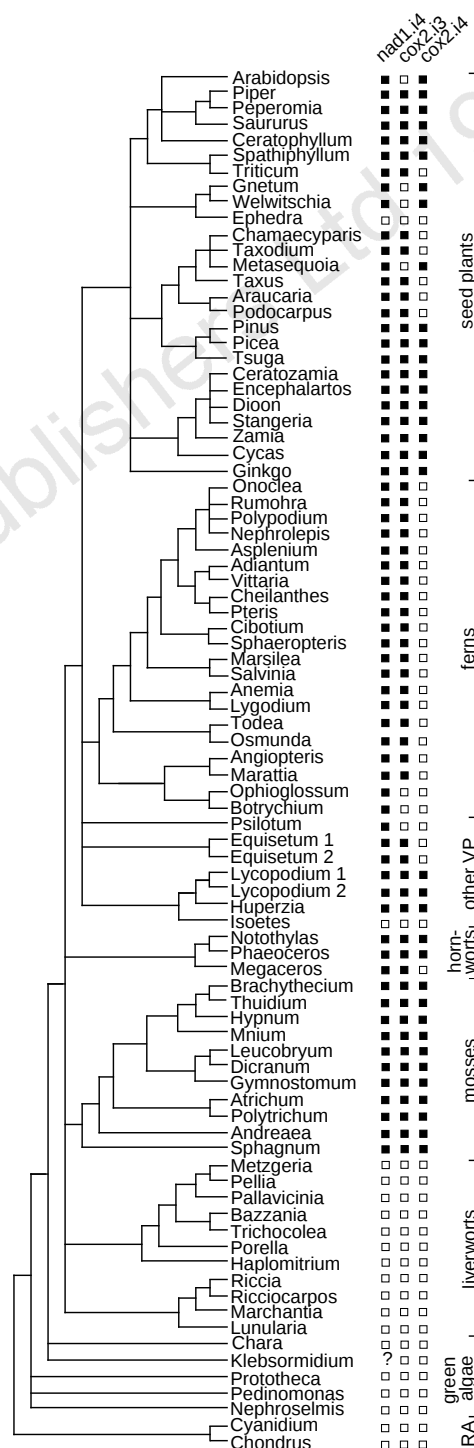


Figure 1 Distribution of three mitochondrial introns among land plants and algae. The tree shows, for 79 diverse land plants and seven algae, a phylogenetic hypothesis based on chloroplast *rbcL* analysis (Y.-L.Q. and J.D.P., unpublished data) and published molecular and morphological studies^{6,7,21–23}. Filled squares indicate intron presence; open squares indicate intron absence. Intron data for land plants were obtained from hybridizations (see Fig. 2), with corroboration by DNA sequencing for selective taxa (Fig. 3); algal data are from sequencing (Fig. 3 and Methods). VP, vascular plants; RA, red algae.

* Present addresses: Institute of Systematic Botany, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland (Y.-L.Q.); Institute of Cellular and Molecular Biology, University of Texas, Austin, Texas 78712, USA (J.C.C.).

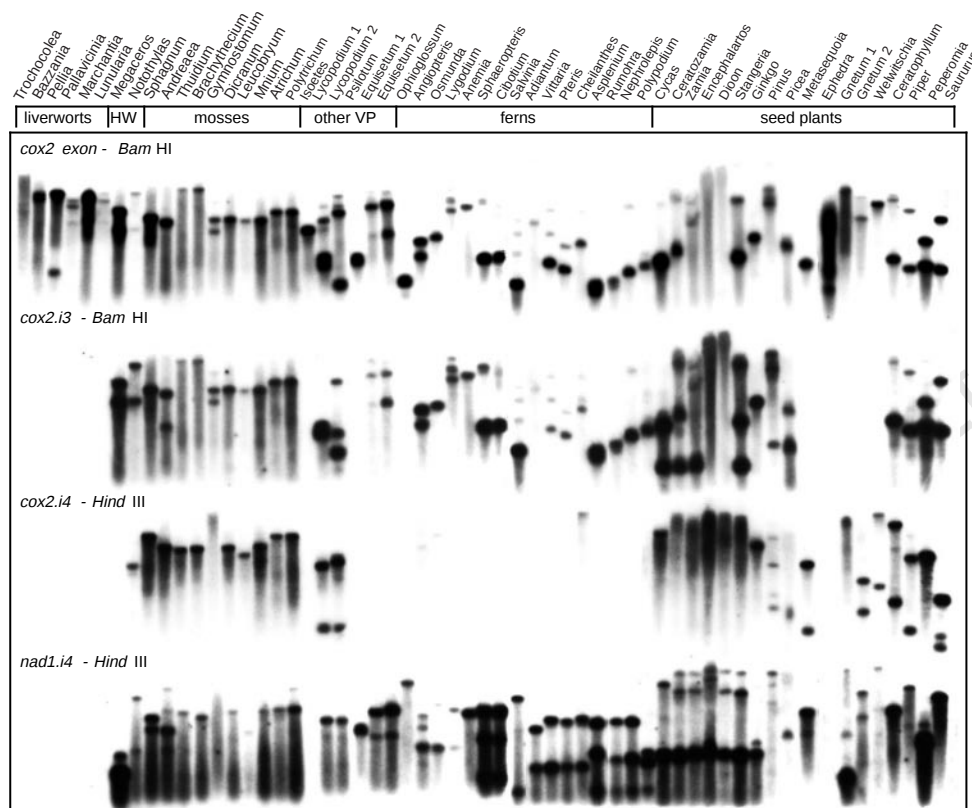


Figure 2 Southern blot hybridizations showing presence or absence of three mitochondrial introns among 58 diverse land plants out of 352 examined in this

study. DNAs from the indicated plants were digested with the indicated enzymes and hybridized with the indicated probes. HW, hornworts; VP, vascular plants.

and exon probes from the same gene co-hybridized to at least one restriction fragment in all but three cases, we inferred that the intron hybridizations represent the presence of a homologous intron in the same gene across taxa (compare the top two panels in Fig. 2, for example; the three exceptions involve the weak hybridization of *cox2.i4* to DNAs from *Cheilanthes*, *Angiopteris* and *Osmunda*; amplification of *cox2* from these taxa by polymerase chain reaction (PCR) confirmed the absence of this intron).

These Southern blot inferences on intron presence/absence were confirmed for 11 key taxa—nine bryophytes and two charophycean green algae—by nucleotide sequencing (Fig. 3). Like the liverworts, the two charophycean green algae lack all three introns (*Klebsormidium* was not examined for *nad1.i4*), as do the three examined non-charophycean green algae, both examined red algae, and all of the many non-plant/non-algal mitochondrial genomes whose *cox2* and *nad1* genes have been sequenced. The introns in the newly sequenced genes from hornworts and mosses are located at identical positions and are highly conserved in nucleotide sequence relative to those in previously characterized angiosperm genes (Fig. 3). For example, the *nad1.i4*, *cox2.i3* and *cox2.i4* introns from the hornwort *Notothylas* are 78, 70 and 65% identical (gaps excluded), respectively, to those of the angiosperms *Arabidopsis* or *Daucus* over the first 420–500 nucleotides of their aligned lengths. These data, together with the widespread presence of the three introns across all major land-plant groups except liverworts, indicate orthology of these introns across land plants; that is, we believe it unlikely that, for example, the hornworts and mosses acquired their introns independently of each other and the vascular plants.

The simplest and most direct interpretation of these data is that all three introns were acquired once during land-plant evolution, in a common ancestor of all land plants exclusive of liverworts. Thus, these data provide evidence in favour of previous suggestions^{5–7} that

the liverworts are the earliest branch of land plants. These results can be reconciled with alternative ideas, for example that hornworts are the earliest branch of land plants or that bryophytes are monophyletic^{8–11}, only by postulating the earlier gain of all three introns, in a common ancestor of all land plants, followed by their complete elimination from all liverworts. The observed occasional loss of each intron (Fig. 1; phylogenetic treatment of these losses will be presented elsewhere) makes this hypothesis tenable, but for the following reasons we consider it unlikely. First, it is clearly less parsimonious. Second, intron losses are rare, with most examined taxa retaining two or all three of the introns; among the 352 land plants examined, only two, clearly derived taxa (*Ephedra* and *Isoetes*) have lost all three introns. Third, if the liverworts are a paraphyletic assemblage of thalloid and leafy liverworts, as suggested by some data^{5,7,13}, then all three introns would have to have been lost at least twice during liverwort evolution. Fourth, the liverwort introns must have all been lost early and coincidentally, at the base of the entire group or its multiple paraphyletic groups. Finally, the presence of numerous (32) mitochondrial introns in the liverwort *Marchantia*¹⁴ suggests that liverworts have not experienced unusual pressures to lose introns.

Identification of liverworts as the earliest lineage(s) of land plants has important implications. This result confirms earlier suggestions of taxonomic bias in the early megafossil record, which is dominated by vascular plants². It means liverworts will serve as valuable models to study such important adaptations to life on land as the multicellular sporophyte (that is, the embryo), cuticle, archegonium and antheridium⁶, as they are the first plants to possess these features. Liverworts should also serve as an excellent outgroup to understand the evolution of stomata and of sporangial columellae, and the change from a gametophyte- to a sporophyte-dominated life cycle. Similarly, the use of liverworts as outgroups should aid molecular attempts to resolve other aspects of land-plant phylogeny,

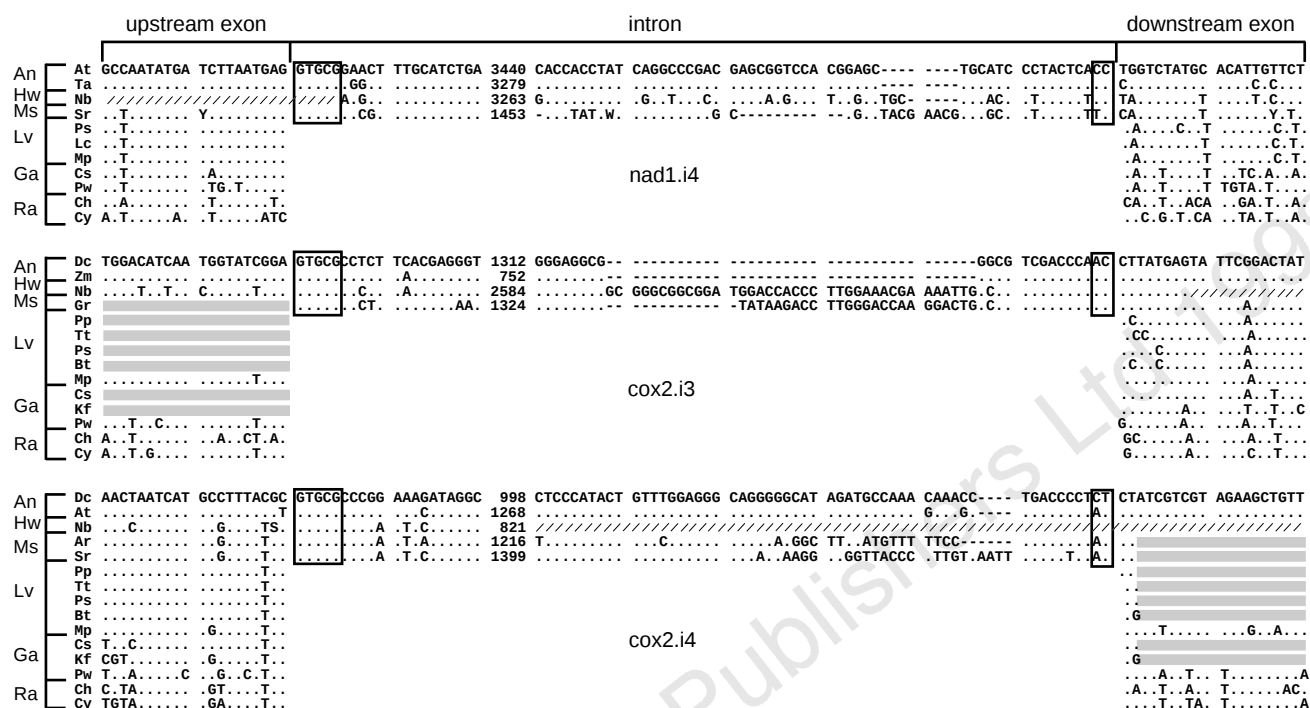


Figure 3 Alignments of exon/intron boundaries for selected intron-containing and intron-lacking land plants and algae. Dots indicate identities to each top sequence; dashes indicate gaps. Numbers indicate lengths of unshown intron sequences. Shaded regions indicate PCR priming sites. Slash marks indicate regions not contained on genomic clones from *Notothylas*. Boxed sequences

as the currently used outgroup (charophycean algae) tends to present a relatively long branch, which is liable to produce long-branch-attraction artefacts^{13,18}. Finally, these findings reinforce the idea that land plants are monophyletic, in that hornworts, the one group sometimes proposed to have a separate origin from the rest of land plants¹⁹, share these three derived introns with all non-liverwort land plants. Although these results raise the possibility that liverworts may have a separate green algal origin from other land plants, we consider this unlikely on the basis of the many derived features shared across all embryophytes¹. □

Methods

Bryophytes were cleaned under a dissecting microscope to avoid contamination. All plant material is vouchered. A list of the 352 plants examined in this study is available at <http://www.bio.indiana.edu/~palmerlab>. Total cellular DNA was prepared by a modified cetyltrimethylammonium bromide method²⁰ and further purified by CsCl/ethidium bromide centrifugation. Southern transfers used Immobilon nylon membranes (Millipore). ³²P-labelled probes were prepared by random priming of PCR fragments internal to: (1) the *cox2* coding sequence (639 base pair (bp) fragment from soybean; GenBank accession number X04825); (2) *cox2.i3* (584 bp from maize; J01425); (3) *cox2.i4* (1198 bp from radish; AF06387); (4) *nad1.i4* (411 bp from *Oenothera*; M63034); and (5) the *nad1* coding sequence (253 bp from *Oenothera*; M63034). Hybridizations were done at 60 °C for 18 h in 5× SSC, 50 mM Tris (pH 8.0), 0.1% SDS, 10 mM EDTA, and 2× Denhardt's solution. Filters were washed twice for 30 min at 60 °C in 2× SSC, 0.5% SDS.

The regions of *cox2* and *nad1* that flanked and included one or more of the three introns under investigation were isolated from 11 bryophytes and green algae by, with one exception, PCR amplification. PCR products were cloned using a TA-cloning kit (Invitrogen). For *Notothylas*, part or all of the three introns and some flanking exonic sequence were isolated by cloning restriction fragments that contained each region, as identified by Southern hybridization using heterologous intron probes. Clones were sequenced on both strands by an automated DNA sequencer (Li-Cor). Sequence data for those algae listed in Fig. 1 are from GenBank (*Chondrus*, *Cyanidium* and *Prototheca*), the unpub-

lished sequences of G. Burger, M. W. Gray and B. F. Lang (*Pedinomonas*), the unpublished sequences of C. Lemieux and M. Turmel (*Nephroselmis*), and this study (*Chara* and *Klebsormidium*). Taxon abbreviations for Fig. 3: Ar, *Andreaea rothii*; At, *Arabidopsis thaliana*; Bt, *Bazzania trilobata*; Cs, *Chara* sp.; Ch, *Chondrus crispus*; Cy, *Cyanidium caldarium*; Dc, *Daucus carota*; Gr, *Gymnostomum rucurvirostrum*; Kf, *Klebsormidium fluitans*; Lc, *Lunularia cruciata*; Mp, *Marchantia polymorpha*; Nb, *Notothylas breutelii*; Ps, *Pallavicinia* sp.; Pp, *Porella pinnata*; Pw, *Prototheca wickerhamii*; Sr, *Sphagnum recurvum*; Tt, *Trichocolea tomentella*; Ta, *Triticum aestivum*; Zm, *Zea mays*. All sequences shown in Fig. 3 are newly determined in this study (GenBank accession numbers AF068930–AF068945) except for those of At, Ch, Cy, Dc, Mp, Pw, Ta and Zm (which are, respectively: Y08501, Z47547, Z48930, X63625, M68929, U02930, X57965, X57966, J01425).

Received 3 March 1998; accepted 15 June 1998.

1. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. *Nature* **389**, 33–39 (1997).
2. Gray, J. Major paleozoic land plant evolutionary bio-events. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **104**, 153–169 (1993).
3. Edwards, D., Duckett, J. G. & Richardson, J. B. Hepatic characters in the earliest land plants. *Nature* **374**, 635–636 (1995).
4. Taylor, W. A. Spores in earliest land plants. *Nature* **373**, 391–392 (1995).
5. Mishler, B. D. *et al.* Phylogenetic relationships of the "green algae" and "bryophytes". *Ann. Mo. Bot. Gard.* **81**, 451–483 (1994).
6. Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants: a Cladistic Study* (Smithsonian Institution Press, Washington DC, 1997).
7. Lewis, L. A., Mishler, B. D. & Vilgals, R. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Mol. Phylog. Evol.* **7**, 377–393 (1997).
8. Hedderston, T. A., Chapman, R. L. & Rootes, W. L. Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. *Plant Syst. Evol.* **200**, 213–224 (1996).
9. Malek, L., Lüttig, K., Hiesel, R., Brennicke, A. & Knoop, V. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J.* **15**, 1403–1411 (1996).
10. Graham, L. E. *Origin of Land Plants* (John Wiley, New York, 1993).
11. Garbary, D. J., Renzaglia, K. S. & Duckett, J. G. The phylogeny of land plants: a cladistic analysis based on male gametogenesis. *Plant Syst. Evol.* **188**, 237–269 (1993).
12. Manhart, J. R. Phylogenetic analysis of green plant *rbcL* sequences. *Mol. Phylog. Evol.* **3**, 114–127 (1994).
13. Capesius, I. & Bopp, M. New classification of liverworts based on molecular and morphological data. *Plant. Syst. Evol.* **207**, 87–97 (1997).
14. Oda, K. *et al.* Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. *J. Mol. Biol.* **223**, 1–7 (1992).

15. Nugent, J. M. & Palmer, J. D. In *Plant Mitochondria* (eds Brennicke, A. & Kück, U.), 163–170 (VCH, Weinheim, 1993).
16. Wolfe, K. H., Li, W.-H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84**, 9054–9058 (1987).
17. Palmer, J. D. & Herbon, L. A. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **28**, 87–97 (1988).
18. Samigullin, T. H. *et al.* Sequences of rDNA internal transcribed spacers from the chloroplast DNA of 26 bryophytes: properties and phylogenetic utilities. *FEBS Lett.* **422**, 47–51 (1998).
19. Sluiman, H. J. A cladistic evaluation of the lower and higher green plants. *Plant Syst. Evol.* **149**, 217–232 (1985).
20. Doyle, J. J. & Doyle, J. S. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
21. Doyle, J. A. & Donoghue, M. J. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Bot. Rev.* **52**, 321–431 (1986).
22. Pryer, K. M., Smith, A. R. & Skog, J. E. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. *Am. Fern J.* **85**, 205–282 (1995).
23. Chaw, S.-M., Zharkikh, A., Sung, H.-M., Lau, T.-C. & Li, W.-H. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Mol. Biol. Evol.* **14**, 56–68 (1997).

Acknowledgements. We thank K. Adams, G. J. Gastony and P. Kuhlman for critical reading of the manuscript, R. C. Brown, B. Crandall-Stotler, G. J. Gastony, B. D. Mishler, K. S. Renzaglia, J. Shaw, H. J. Sluiman, A. R. Smith, S. H. Strauss, D. Waters and J. A. Wheeler for plant material, F. Dong and K. G. Wilson for probes, and G. Burger, M. W. Gray, B. F. Lang, C. Lemieux and M. Turmel for unpublished data. This work was supported by grants to Y.Q. and J.D.P. from the NIH.

Correspondence and requests for materials should be addressed to J.D.P. (e-mail: jpalmer@bio.indiana.edu).

Noise and determinism in synchronized sheep dynamics

B. T. Grenfell*, K. Wilson†, B. F. Finkenstädt*, T. N. Coulson‡, S. Murray§, S. D. Albon||, J. M. Pemberton¶, T. H. Clutton-Brock* & M. J. Crawley#

* Zoology Department, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

† Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, UK

‡ Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

§ Scottish Natural Heritage, Clachcaran, Stilligarry, South Uist, Western Isles HS8 5RS, UK

|| Institute of Terrestrial Ecology, Hill of Brathens, Glassel, Banchory AB31 4BY, UK

¶ Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

Imperial College, Silwood Park, Ascot SL5 7PY, UK

A major debate in ecology concerns the relative importance of intrinsic factors and extrinsic environmental variations in determining population size fluctuations^{1–6}. Spatial correlation of fluctuations in different populations caused by synchronous environmental shocks^{2,7,8} is a powerful tool for quantifying the impact of environmental variations on population dynamics^{8,9}. However, interpretation of synchrony is often complicated by migration between populations^{8,10}. Here we address this issue by using time series from sheep populations on two islands in the St Kilda archipelago^{11–13}. Fluctuations in the sizes of the two populations are remarkably synchronized over a 40-year period. A nonlinear time-series model shows that a high and frequent degree of environmental correlation is required to achieve this level of synchrony. The model indicates that if there were less environmental correlation, population dynamics would be much less synchronous than is observed. This is because of a threshold effect that is dependent on population size; the threshold magnifies random differences between populations. A refined model shows that part of the required environmental synchronicity can be accounted for by large-scale weather variations. These results underline the importance of understanding the interaction between intrinsic and extrinsic influences on population dynamics¹⁴.

Much ecological debate has focused on the interaction between noise and nonlinear dynamics in generating population cycles^{5,6,13,15–18} and patterns of spatial synchrony^{8,10,19}. Here we use the unusual circumstance of long parallel time series from close—but completely isolated—populations as a tool with which to explore these issues.

Feral sheep populations on islands in the St Kilda archipelago have been monitored since 1955 (ref. 11) (Fig. 1). The most detailed series comprise continuous annual records for Soay sheep on the main island, Hirta, and 18 years of counts for Blackface sheep on Boreray; the islands are 3.5 km apart.

Both series show irregular population fluctuations, reflecting repeated mass mortalities of sheep^{12,20} (Fig. 1a). Counts of sheep on the two islands are highly correlated (Fig. 1b) for the logged series. Pearson's $r = 0.685$ (95% bootstrap confidence limits are 0.447–0.838). As the two populations are completely separate, this synchrony indicates a high correlation in extrinsic environmental

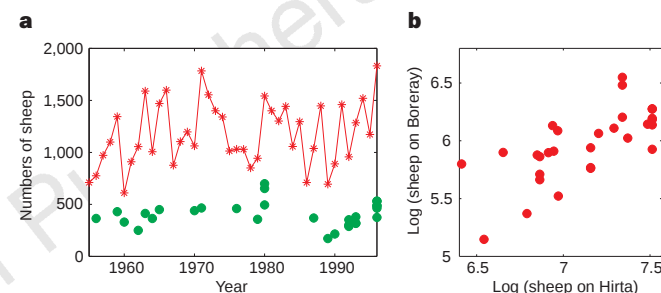


Figure 1 Feral sheep populations on Hirta and Boreray. **a**, Time series of total sheep counts from Hirta (red)^{10,19} and Boreray (green). Boreray counts were obtained either on the island itself or from a circumnavigating boat. Both methods allow most of the island to be observed and produce similar figures. The Boreray counts are replicated in some years, showing that the estimates are consistent. **b**, Scatter plot of the two island time series (logged). To allow easy comparison with the models, we calculated the correlation coefficient after replacing replicate Boreray counts for a given year with their maximum.

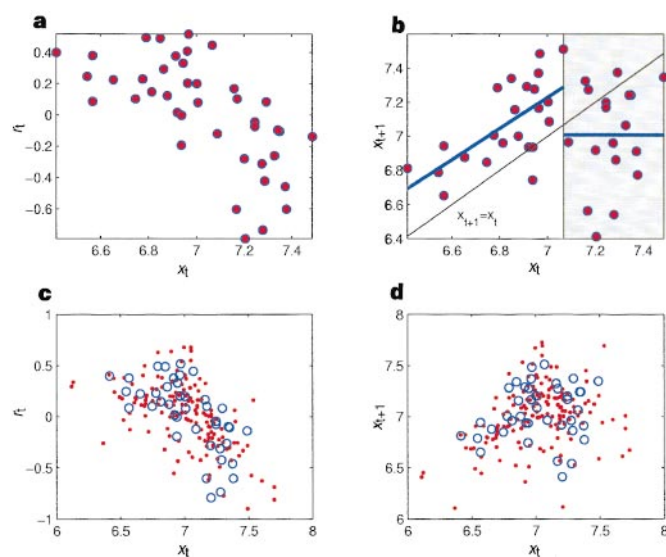


Figure 2 Modelling the Hirta time series. **a**, Plot of annual population growth rate, $r_t = x_{t+1} - x_t$, against log population size, x_t . **b**, Fit of univariate SETAR model (see Table 1a) to the scatter plot of x_{t+1} against x_t . The shaded area shows the regime above the population threshold, $x_t = C = 7.066$; blue lines show the best-fit model; and the diagonal black line is at $x_{t+1} = x_t$. **c**, A comparison of the observed r_t versus x_t plot (blue, open circles) with 150 iterates of the best-fit model with added noise (red dots), as defined in Table 1; a transient of 250 years was run off before recording the points. **d**, The same comparison as that shown in **c**, but plotted as x_{t+1} against x_t .

Table 1 Fits of threshold autoregressive models to the Hirta Soay sheep time series

	Estimate	Standard error	<i>t</i>	<i>P</i>
a Basic threshold model				
Regime 0: below-fit threshold, $C = 7.066$				
a_0	0.848	1.564	0.542	0.3*
b_0	0.912	0.229†	3.98	<10 ⁻⁵ *
σ_0	0.183			
Regime 1: above threshold				
A_1	7.01	0.069	101	<10 ⁻⁵
b_1	0	—	—	—
σ_1	0.293			
b Threshold model with weather covariates				
Regime 0: below best-fit threshold, $C = 7.17$				
a_0	0.0217	1.66	0.013	0.495
b_0	0.811	0.209	3.883	0.0005
c_0	-0.0073	0.0004	-1.769	0.05
d_0	0.214	0.052	4.08	0.0003
σ_0	0.149			
Regime 1: above threshold				
a_1	7.174	0.0955	75	<10 ⁻⁵
b_1	0	—	—	—
c_1	-0.00157	0.0007	-2.208	0.024
d_1	0	—	—	—
σ_1	0.246			

Parameter estimates and associated significance levels are shown for the best-fit models; a zero coefficient indicates that the relevant term is not significant enough to be included. **a**, Estimates for the model without weather covariates (equation (1)). **b**, The model with weather covariates (equation (2)). We included two significant weather variables g , the number of hours of March gales, and h , the mean April temperature (precipitation did not account for any significant variation).

* Probabilities are only approximate levels of significant differences from zero; the AIC is the criterion used for establishing the minimal model.

† b_0 is not significantly different from 1.

influences on population fluctuations⁷. We studied the causes of population synchronization using a simple model for the relative influence on sheep population dynamics of forces dependent on extrinsic noise and intrinsic density.

We analysed the data as log (population numbers), $x_t = \log(N_t)$ (see Methods). Figure 2a shows the basic pattern of density dependence in the Hirta population by a standard plot of log (population growth rate), $r_t = x_{t+1} - x_t$, against log (population numbers), x_t (ref. 2). The figure shows that growth rate declines as population size increases, as a result of intraspecific competition for food^{11,20}. There is also marked variation around this trend and the variance is greater at high population density.

The general approach to modelling this interaction between noise and density dependence is to express x_{t+1} as a function (f) of previous population sizes, as follows: $x_{t+1} = f(x_t, x_{t-1}, x_{t-2}, \dots) + \epsilon_{t+1}$ (refs 21–23), where ϵ represents the additive variation around the relationship. In some cases, f reflects a nonlinear relationship between current and past population sizes²³. To allow for this possibility, we fit a nonlinear self-exciting threshold autoregressive (SETAR) model^{23–25} to the Hirta time series. This method approaches the problem of nonlinearity in time-series models by estimating the optimal piecewise linear model for the function f (Table 1 and Fig. 2b).

The analysis (Table 1a) shows that only time lags of up to one year account for sufficient variation to be included in the model. The best-fit model

$$\begin{aligned} x_{t+1} &= a_0 + b_0 x_t + \epsilon_{0,t+1} & x_t &\leq C \\ x_{t+1} &= a_1 + \epsilon_{1,t+1} & x_t &> C \end{aligned} \quad (1)$$

displays evidence of nonlinearity, captured by a change in dynamics at a population threshold, C ($\exp(C) = 1,172$ individuals); this is seen in the plot of x_{t+1} against x_t (Fig. 2b). Below C , the dynamics follow a simple recurrence relationship between x_{t+1} and x_t , modified by additive normally distributed noise ($\epsilon_{0,t+1} \sim N(0, \sigma_0^2)$). The intercept, a_0 , controls the mean density-independent population growth rate. The estimated slope, b_0 , is not significantly different from unity (Table 1a), indicating that there is no evidence of density-dependent constraints on population growth at low

numbers²¹. In contrast, above the threshold, x_{t+1} is independent of x_t and highly variable (Fig. 2b).

The model shows that there is a noisy exponential increase in population size when numbers are low following a population crash. At high densities, the population can increase, remain constant or fall, depending on environmental conditions. Simulation of the model with noise captures the essential features of both the growth-rate pattern (Fig. 2c) and the map of x_{t+1} versus x_t (Fig. 2d).

What does this analysis tell us about the high level of observed correlation between the Hirta and Boreray populations? Assuming that the Hirta model applies to both islands, we can use it to estimate the level of correlation in environmental noise required to generate the observed synchrony in population fluctuations. Specifically, we assessed how the nonlinear change in dynamics across the population threshold, C , affects the relationship between environmental correlation and population synchrony. An appropriate null model is provided by the Moran effect⁷, which applies to the synchrony of isolated populations described by a common linear autoregressive regime. Moran's theorem² states that the asymptotic correlation of populations following identical linear autoregressive models will equal the correlation between random environmental perturbations. In a single regime with no threshold, the expected environmental correlation would therefore equal the observed correlation between populations, $r = 0.685$, as shown in Fig. 3.

In fact, Fig. 3a shows that much higher levels of noise correlation ($r > 0.9$, on average) are needed to generate the observed correlation between sheep populations on the two islands. This is because of the density-dependent nonlinearity in the system², which we capture by the threshold. Fluctuations both above and below the

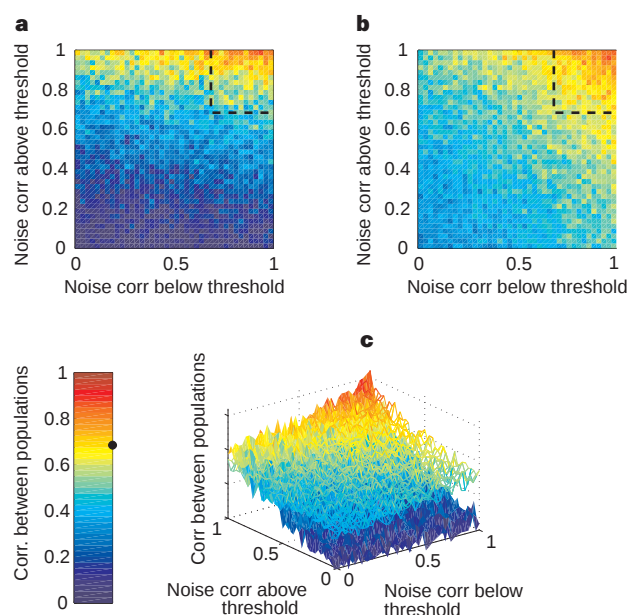


Figure 3 Simulations of the observed correlation in sheep counts between islands. Results are shown as contour maps of population correlation, as a function of inter-island correlation in environmental noise below and above the population threshold. Population correlation is colour coded as shown in the colour key, which marks the observed population correlation ($r = 0.685$) as a black dot. **a**, Correlations from the basic SETAR model, equation (1); the dashed lines show the expected environmental correlation ($r = 0.685$) required to explain the island synchrony, based on Moran's theorem. **b**, Correlation contours, as in **a**, but using the model with added weather covariates (equation (2)). **c**, Island correlations for the weather covariate model, as in **b**, but plotted as a (transparent) surface. The equivalent surface for the basic SETAR model (**a**) is shown as the underlying solid surface.

threshold need to be highly synchronized to generate the observed population correlation. The asymmetry of Fig. 3a also indicates that environmental synchrony above the threshold is the more powerful correlating force.

The correlation analysis shows that the populations are influenced by frequent extrinsic environmental variations at all population levels. We studied underlying mechanisms by including observed large-scale meteorological covariates (monthly wind, rain and temperature) in the model (see Table 1b). The best-fit model:

$$\begin{aligned} x_{t+1} &= a_0 + b_0 x_t + c_0 g_{t+1} + d_0 h_{t+1} + \epsilon_{0,t+1} & x_t \leq C \\ x_{t+1} &= a_1 + c_1 g_{t+1} + \epsilon_{1,t+1} & x_t > C \end{aligned} \quad (2)$$

has the same basic structure as equation (1), but now includes two extra terms. First, it includes a negative effect of March gales (hours of gales, g_{t+1}) on population growth rate, both above and below the threshold: gales are likely to increase the death rate of sheep, which are at their lowest physiological ebb at this time²⁰. Second, a positive effect of April temperature, h_{t+1} , on population growth rate below the threshold is included; presumably temperature affects the timing of rapid grass growth. The parameters c_0 , c_1 and d_0 , d_1 control the strength of gale and temperature effects, respectively.

The new model allows us to assess how synchrony in weather affects the inter-island population correlation. We assume that the gales and temperature are the same for the two islands. How much 'extra' correlation do we need in the remaining unexplained variability above and below the threshold to generate the observed level of population correlation? To determine this, we repeated the correlation analysis of Fig. 3a using the model with added weather covariates (Fig. 3b, c). The inclusion of common weather effects can account for part of the inter-island population correlation—the 'unexplained' residual environmental correlation required to the population synchrony is reduced from >0.9 to around 0.7 (Fig. 3b). The common weather pattern generates an inter-island population correlation of about 0.3 when the unexplained residual noise is uncorrelated (that is, at the three-dimensional origin of Fig. 3c).

This analysis underlines that the observed inter-island population synchrony is remarkably high. We required an average environmental correlation of over 0.9—well above the prediction of the Moran effect—to achieve the observed population correlation of ~0.7. This is because of the density-dependent nonlinearity in the system². Specifically, the two sheep populations must both be sufficiently entrained by the same environmental conditions to cross the population threshold, C , in the same year if they are to remain highly synchronized. This result has general implications for the application of models for environmental entrainment of spatial dynamics^{8,10,19} to systems that exhibit threshold behaviour.

Analysis of the inter-island correlation also provides a powerful tool for dissecting the internal dynamics of the populations. First, the SETAR model (equation (1)) extends the conclusions of previous studies of the Hirta population^{13,20}. There is a threshold sheep density that is determined by food availability; above this density the population can undergo spectacular crashes, which are consistent with strong, and sharply focused, overcompensatory density dependence¹³. However, populations above the threshold need not necessarily crash. At high population density, even though the average density-independent growth rate is negative (Fig. 2a), the variance in population growth rate is high enough that the population size can sometimes increase. Essentially the same dynamical pattern is apparent in the Boreray series.

Second, the inclusion of meteorological covariates in the model explains part of the interaction between density-dependent and -independent forces. The population takes two years after a population crash to rise again to high levels²⁰. If this rise above the threshold coincides with severe spring gales, they increase the chance of another severe crash. The probability and degree of overcompensatory density dependence therefore depend on the weather.

Finally, the portion of inter-island correlation that remains unexplained by the available weather covariates (Fig. 3d) shows that there is a good deal more to understand about the dynamics of crashes in mechanistic terms. Specifically, it seems likely that some combination of food shortage¹², parasitism^{26,27} and the timing of bad weather stresses the sheep enough to precipitate crashes. In high-density years, the system appears to be exquisitely sensitive to this combination of density-dependent and -independent factors.

Our results underline the importance of detailed long-term studies for teasing out the interaction of deterministic and stochastic influences on population fluctuations¹⁴. Threshold models with meteorological covariates²⁴ provide a useful method with which to analyse these issues. □

Methods

Fitting the threshold autoregressive models. This general approach to modelling nonlinearity in time series²⁴ partitions the process into regimes that follow linear autoregressive models. Estimation is relatively straightforward, as the linear theory applies to the subregimes²⁵. The method is particularly useful for the sheep series as it allows, first, for changes in noise variance with population size (by permitting different noise levels for each regime) and, second, the straightforward addition of weather covariates (Table 1). We based the model on log (population number), as this allows naturally for the possibility of density dependence within each regime, as indicated by a previous model of the system¹³, which showed threshold-like behaviour for the 1985–91 data. In fact, a threshold model based on absolute population abundance (N_t) generates similar qualitative results to the models used here.

An important issue in model selection is the order (maximum number of lags) in each regime. We estimated this up to a lag of five years, by minimizing the Akaike information criterion (AIC)^{24,28}. We also used the AIC to identify the best-fit threshold model, allowing for degrees of freedom. The best-fit SETAR model is of order 1, with five parameters (AIC = -109.97, compared to -107.57 for the best linear autoregressive (0) model with one parameter). The fit indicates a significant discontinuity in the observed dynamics across the threshold (captured by the break in the blue lines in Fig. 2b). Forcing a continuous unimodal deterministic skeleton (either a broken line model, or a low-order polynomial) gives qualitatively the same island correlation results as those shown in Fig. 3. However, such models (not shown) do not capture the observed threshold in both noise and determinism as well as does equation (1).

Weather covariates. We tested a range of weather variables for inclusion in the model. On Hirta most mortality occurs at the end of winter in March and April when many animals can die of starvation^{11,12}. Research on other breeds of sheep has shown that extreme weather conditions at times close to lambing can cause perinatal mortality²⁹ and that poor weather can decrease the proportion of time an individual spends feeding³⁰. Consequently we considered mean monthly temperature, mean monthly precipitation and the total number of hours with wind speeds of over 34 knots during the months of March and April as possible causes of the observed correlations between islands. Weather data were provided by the UK Meteorological Office and come from the closest meteorological station, on the island of Benbecula (Outer Hebrides), 80 km east of St Kilda.

Simulations of the island correlation. We calculated each point in the 40 × 40 contour maps (Fig. 3) from replicate pairs of simulations of the specified model. The second simulation had a different series of random shocks from the first, adjusted to give the correlation between the two noise series above and below the threshold, C , specified by the axes of the plot. Correlations between the (logged) series were calculated for 800 years, after a transient of 250 years. We calculated the simulated weather covariates shown in Fig. 3c by bootstrapping (resampling with replacement) the observed series, to generate long, simulated time series of March gales and April temperatures. We sampled the weather data as annual (gale, temperature) pairs.

Received 10 October 1997; accepted 2 June 1998.

1. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, NJ, 1973).
2. Royama, T. *Analytical Population Dynamics* (Chapman and Hall, London, 1992).
3. Stenseth, N. C., Bjørnstad, O. N. & Saitoh, T. A gradient from stable to cyclic populations of *Clethrionomys rufocanus* in Hokkaido, Japan. *Proc. R. Soc. Lond. B* **263**, 1117–1126 (1996).
4. Sugihara, G. Ecology—from out of the blue. *Nature* **378**, 559–560 (1995).

5. Ellner, S. & Turchin, P. Chaos in a noisy world—new methods and evidence from time-series analysis. *Am. Nat.* **145**, 343–375 (1995).
6. Leirs, H. *et al.* Stochastic seasonality and nonlinear density-dependent factors regulate population size in an African rodent. *Nature* **389**, 176–180 (1997).
7. Moran, P. A. P. The statistical analysis of the Canadian lynx cycle: II synchronisation and meteorology. *Aust. J. Zool.* **1**, 291–298 (1953).
8. Ranta, E., Kaitala, V., Lindström, J. & Helle, E. The Moran effect and synchrony in population dynamics. *Oikos* **78**, 136–142 (1997).
9. Hill, J. K., Thomas, C. D. & Lewis, O. T. Effects of habitat patch size and isolation on dispersal by *Hesperia comma* butterflies: implications for metapopulation structure. *J. Anim. Ecol.* **65**, 725–735 (1996).
10. Bjørnstad, O. N., Falck, W. & Stenseth, N. C. Geographic gradient in small rodent density fluctuations—a statistical modeling approach. *Proc. R. Soc. Lond. B* **262**, 127–133 (1995).
11. Jewell, P. A., Milner, C. & Morton-Boyd, J. *Island Survivors* (Athlone, London, 1974).
12. Clutton-Brock, T. H., Price, O. F., Albon, S. D. & Jewell, P. A. Persistent instability and population regulation in Soay sheep. *J. Anim. Ecol.* **60**, 593–608 (1991).
13. Grenfell, B. T., Price, O. F., Albon, S. D. & Clutton-Brock, T. H. Overcompensation and population cycles in an ungulate. *Nature* **355**, 823–826 (1992).
14. Sugihara, G. Red/blue chaotic power spectra. *Nature* **381**, 199–199 (1996).
15. Grenfell, B. T. Chance and chaos in measles dynamics. *J. R. Stat. Soc. B* **54**, 383–398 (1992).
16. Falck, W., Bjørnstad, O. N. & Stenseth, N. C. Voles and lemmings—chaos and uncertainty in fluctuating populations. *Proc. R. Soc. Lond. B* **262**, 363–370 (1995).
17. Stenseth, N. C. Snowshoe hare populations—squeezed from below and above. *Science* **269**, 1061–1062 (1995).
18. Higgins, K., Hastings, A., Sarvela, J. N. & Botsford, L. W. Stochastic dynamics and deterministic skeletons: population behavior of Dungeness crab. *Science* **276**, 1431–1435 (1997).
19. Ranta, E., Kaitala, V., Lindström, J. & Linden, H. Synchrony in population dynamics. *Proc. R. Soc. Lond. B* **262**, 113–118 (1995).
20. Clutton-Brock, T. H. *et al.* Stability and instability in ungulate populations: an empirical analysis. *Am. Nat.* **149**, 195–219 (1997).
21. Hassell, M. P., Lawton, J. H. & May, R. M. Patterns of dynamical behaviour in single-species populations. *J. Anim. Ecol.* **45**, 471–486 (1976).
22. Turchin, P. Nonlinear time-series modeling of vole population fluctuations. *Res. Pop. Ecol.* **38**, 121–132 (1996).
23. Stenseth, N. C., Falck, W., Bjørnstad, O. N. & Krebs, C. J. Population regulation in snowshoe hare and Canadian lynx: asymmetric food web configurations between hare and lynx. *Proc. Natl Acad. Sci. USA* **94**, 5147–5152 (1997).
24. Tong, H. *Non-linear Time Series: a Dynamical Systems Approach* (Oxford Univ. Press, 1990).
25. Tsay, R. S. Testing and modelling threshold autoregressive processes. *J. Am. Stat. Assoc.* **84**, 230–240 (1989).
26. Gulland, F. M. D. Role of nematode infections in mortality of Soay sheep during a population crash on St Kilda. *Parasitology* **105**, 493–503 (1992).
27. Grenfell, B. T., Wilson, K., Isham, V. S., Boyd, H. E. G. & Dietz, K. Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. *Parasitology* **111**, S135–S151 (1995).
28. Framstad, E., Stenseth, N. C., Bjørnstad, O. N. & Falck, W. Limit cycles in Norwegian lemmings: tensions between phase-dependence and density-dependence. *Proc. R. Soc. Lond. B* **264**, 31–38 (1997).
29. Cloete, S. W. P., Van Halderen, A. & Schneider, D. J. Causes of perinatal lamb mortality amongst dromer and SA mutton Merino lambs. *J. S. Afr. Vet. Assoc.* **64**, 121–125 (1993).
30. Champion, R. A., Rutter, S. M., Penning, P. D. & Rook, A. J. Temporal variation in grazing behavior of sheep and the reliability of sampling periods. *Appl. Anim. Behav. Sci.* **42**, 99–108 (1994).

Acknowledgements. We thank Scottish Natural Heritage and the National Trust for Scotland for permission to work on St Kilda, and their staff for assistance, support and encouragement; J. M. Pilkington, A. McColl, A. Robertson and I. R. Stevenson for help with the project; P. Rohani for comments on the manuscript; H. Tong for statistical advice and encouragement; and the Royal Artillery for logistical support on St Kilda. The work was funded by grants from the Natural Environment Research Council (to T.H.C.-B., M.J.C., B.T.G., K.W. and S.D.A.), the Royal Society (to M.J.C., K.W. and S.D.A.) and the Biotechnology and Biological Sciences Research Council (to J.M.P.).

Correspondence and requests for materials should be addressed to B.T.G. (e-mail: bryan@zoo.cam.ac.uk).

Cortical area MT and the perception of stereoscopic depth

Gregory C. DeAngelis*, Bruce G. Cumming† & William T. Newsome*

* Howard Hughes Medical Institute and Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA
† University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK

Stereopsis is the perception of depth based on small positional differences between images formed on the two retinæ (known as binocular disparity). Neurons that respond selectively to binocular disparity were first described three decades ago^{1,2}, and have since been observed in many visual areas of the primate brain, including V1, V2, V3, MT and MST^{3–8}. Although disparity-selective neurons are thought to form the neural substrate for stereopsis, the mere existence of disparity-selective neurons does not

guarantee that they contribute to stereoscopic depth perception. Some disparity-selective neurons may play other roles, such as guiding vergence eye movements^{9,10}. Thus, the roles of different visual areas in stereopsis remain poorly defined. Here we show that visual area MT is important in stereoscopic vision: electrical stimulation of clusters of disparity-selective MT neurons can bias perceptual judgements of depth, and the bias is predictable from the disparity preference of neurons at the stimulation site. These

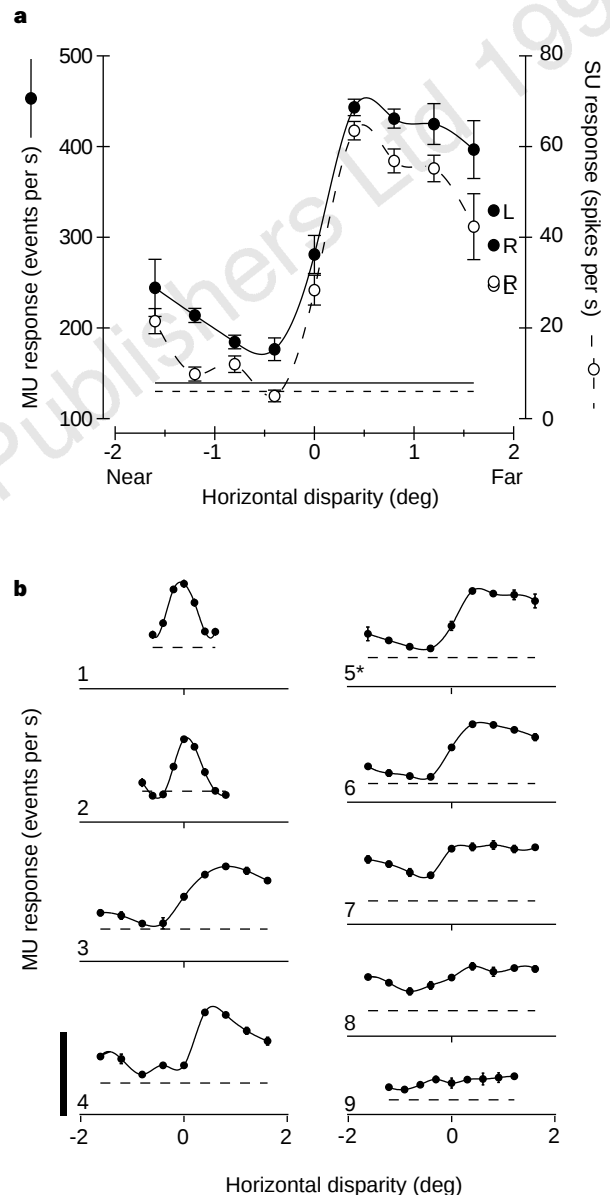


Figure 1 MT neurons are clustered according to disparity selectivity. **a**, Filled circles show multiunit (MU) responses to a drifting random-dot pattern; each datum is the mean of four responses ± 1 s.e. The solid curve is a cubic spline interpolation. Filled circles labelled 'L' and 'R' denote multiunit responses to the same visual stimulus presented monocularly to either the left or the right eye, respectively. The solid horizontal line gives the multiunit response in the absence of a visual stimulus (spontaneous activity). Open circles and the dashed curve show responses of an isolated single unit (SU) recorded simultaneously. The dashed horizontal line gives the spontaneous activity level of the single unit. **b**, Sequence of disparity-tuning curves recorded at 100 μ m intervals along an electrode penetration through MT in monkey S. Standard error bars are generally hidden by the data points. Curves are cubic spline interpolations, and dashed lines represent the spontaneous activity level. Height of scale bar is 400 events per second. Multiunit responses from site 5 (marked by an asterisk) are the same data shown in **a**.

results show that behaviourally relevant signals concerning stereoscopic depth are present in MT.

The middle temporal visual area (MT or V5) is important in motion perception¹¹. Most MT neurons are directionally selective, and these neurons are organized into a system of direction columns such that neurons in a particular column respond optimally to a specific direction of motion¹². Electrical microstimulation of a column of MT neurons can bias perceptual judgements of motion direction towards the direction encoded by the stimulated neurons^{13,14}. We have observed, however, that clusters of neighbouring MT neurons also respond optimally to a common range of preferred binocular disparities¹⁵. Figure 1a, for example, illustrates disparity tuning for a single unit and for a multiunit cluster, measured simultaneously from a single microelectrode. Action potentials from the single unit were excluded from the multiunit response, thus ensuring that the multiunit activity arose from several other nearby single units. The multiunit response is strongly tuned for disparity, indicating that its constituent single units have a similar disparity preference. This conclusion is further supported by the close agreement in shape between the multiunit and single-unit tuning curves; this correspondence is typical of most of our recordings in MT. Roughly half of the multiunit recording sites in MT showed strong disparity tuning (disparity-tuning index > 0.5; see Methods). When multiunit clusters were unselective for disparity, simultaneously recorded single units were generally unselective as well, eliminating the possibility that weak multiunit tuning results from combinations of single units tuned to widely different disparities.

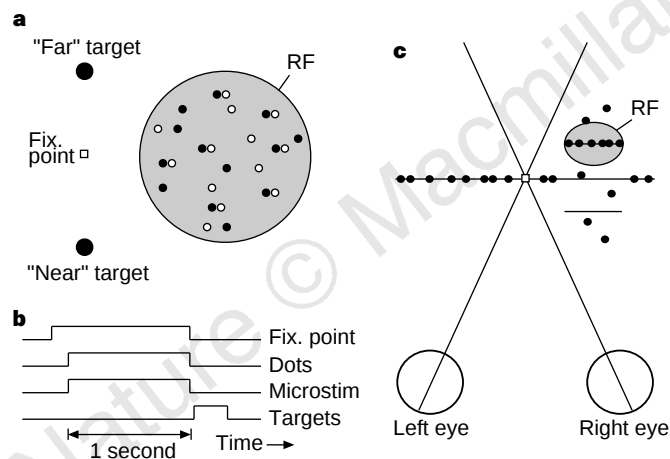


Figure 2 Depth-discrimination task. **a**, Spatial arrangement of fixation point, visual stimulus and response targets. A random-dot pattern was presented within a circular aperture that was roughly the same size as the multiunit receptive field (RF; shaded circle). Filled dots denote the image shown to the left eye, and unfilled dots indicate the image shown to the right eye. Half of the dots are shown paired with a fixed disparity ('signal' dots), and the remaining dots have random horizontal disparities ('noise' dots). **b**, Sequence of trial events. The fixation point appeared first, and the monkey was required to maintain fixation within a $\pm 1.5^\circ$ window throughout the trial. After fixation, the random-dot pattern appeared for 1 second. In half the trials, selected randomly, microstimulation was applied concurrently with the dots. Subsequently, fixation point, visual stimulus and microstimulation were turned off, and two small target disks appeared. The monkey made a saccadic eye movement to the upper target to indicate a far stimulus, or to the lower target to indicate a near stimulus. **c**, Top-down view of the visual stimulus. In each trial, signal dots appeared at one of two fixed disparities, one near and one far (short horizontal line segments). One disparity (far in this example) was chosen to be optimal for the recorded neurons; the other disparity was chosen to be non-optimal. Here, the visual stimulus consists of 50% signal dots and 50% noise dots. Signal and noise dots moved with a velocity (direction and speed) that was optimal for the neurons at the electrode tip. Dots situated along the plane of fixation (long horizontal line) were stationary and provided a zero-disparity background.

Figure 1b shows a sequence of disparity-tuning curves measured at 100- μ m intervals along an oblique electrode penetration through MT. At recording sites 1 and 2, the multiunit activity was well tuned, with a preference for zero disparity. At site 3, there was a transition to broader tuning with a strong preference for far (positive) disparities over near disparities. This pattern was maintained over a span of about 300 μ m (sites 3–6), after which the disparity tuning gradually became weaker, from site 7 to site 9. In other penetrations, regions of poor disparity tuning could span 1 mm or more, and were usually followed by regions of strong tuning. Thus, strongly disparity-tuned neurons are found within discrete patches in MT. The presence of extended regions of similar disparity tuning, such as sites 3–6 in Fig. 1b, is reminiscent of the well-known clustering of MT neurons by preferred direction, which we have exploited in previous microstimulation studies¹³. We therefore tested the role of MT in depth perception by applying electrical stimulation to disparity-tuned sites while two monkeys performed a stereoscopic depth-discrimination task. To activate a cluster of MT neurons with similar tuning, we applied stimulation to sites in which multiunit disparity tuning remained fairly constant over at least 200–300 μ m (for example, sites 3–6 in Fig. 1b).

Figure 2 illustrates the depth-discrimination task. Monkeys viewed a random-dot visual display in which a fraction of the dots carried a consistent depth signal ('signal' dots), whereas the remaining dots were randomly scattered in depth ('noise' dots). In each trial, signal dots were presented at either a near or a far disparity, one of which was chosen to be optimal for the MT neurons recorded at the electrode tip. We chose other parameters of the visual stimulus (location, size, direction and speed of motion) to maximize the multiunit response. The monkey's task was to report seeing near or far depth by making a saccadic eye movement to one of two small visual targets. We varied task difficulty by adjusting the fraction of signal dots in the visual display (per cent binocular correlation). If microstimulation augments the signal carried by the stimulated neurons, and if these neurons provide signals used by the monkey in performing the task, then microstimulation should bias the monkey's choices in favour of the preferred disparity at the stimulation site. Importantly, the monkey was always rewarded for correctly reporting the actual depth of the signal dots, irrespective of the presence or absence of microstimulation. Thus the reward contingencies always encouraged veridical performance.

Figure 3 shows the results from two experiments in which microstimulation produced a bias toward the preferred disparity. Figure 3a depicts the multiunit disparity-tuning curve for a stimulation site that preferred far disparities, and the effect of microstimulation at this site on the monkey's psychophysical judgements is shown in Fig. 3b. At almost every binocular correlation level, the monkey made more 'preferred' decisions in stimulated trials than in non-stimulated trials. The net effect of this preferred disparity bias is a leftward shift of the psychometric function, which we measured as the horizontal offset between sigmoidal curves that were fitted to the data using logistic regression. In this experiment, the shift was equivalent to 22.5% correlated dots, an average-sized effect for monkey R.

Figure 3c shows the disparity-tuning curve for a stimulation site in monkey T that preferred near disparities. Again, microstimulation biased the animal's choices toward the preferred (near) disparity of the stimulated neurons. The leftward shift of the psychometric function was equivalent to 57.6% correlated dots (Fig. 3d), the largest effect that we have observed so far.

Figure 4 summarizes results from 65 microstimulation experiments performed in two monkeys. The scatter plot shows the size of the microstimulation effect plotted against the disparity-tuning index (DTI) of multiunit activity at each stimulation site. Positive values on the ordinate correspond to leftward shifts of the psychometric function, that is, shifts toward the preferred disparity of the stimulated neurons. Filled symbols indicate statistically significant

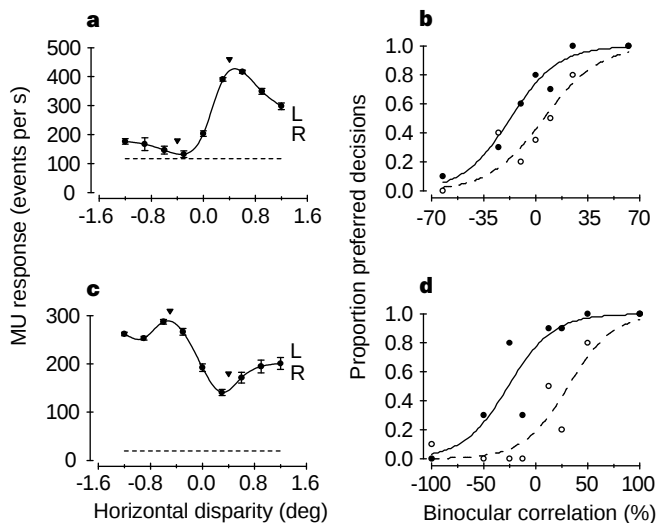


Figure 3 Microstimulation of MT biases depth judgements. **a**, Disparity tuning of multiunit activity at a site in monkey R that prefers far disparities. Arrowheads denote the two disparities used in the discrimination task. L and R denote multiunit responses to the same visual stimulus presented monocularly to either the left or the right eye, respectively. **b**, Effect of microstimulation on depth judgements for the site depicted in **a**. The abscissa gives the percentage of binocularly correlated dots (that is, the percentage of 'signal' dots). Positive values indicate that signal dots were presented at the preferred disparity; negative values correspond to signal dots presented at the non-optimal disparity. The ordinate shows the proportion of decisions made in favour of the preferred disparity (10 trials per point). Filled and open circles correspond to trials with and without microstimulation, respectively. The effect of microstimulation was highly significant (logistic regression, $P < 0.005$). **c**, Disparity tuning of multiunit activity at a near-tuned site in monkey T. **d**, Effect of microstimulation at the site for which tuning is shown in **c**. This effect was also significant ($P < 0.0001$).

effects (logistic regression¹⁶, $P < 0.05$), which occurred in 43 of 65 experiments. Among the significant effects, 42 of 43 were in the direction predicted by the disparity tuning of the multiunit activity; the one exception occurred at a site with weak disparity tuning. In contrast, microstimulation had a significant effect on the slope of the psychometric function in only 4 of 65 experiments. The mean slope ratio (stimulation versus no stimulation) was 0.97, which is not significantly different from 1.0 (t -test, $t = -0.68$, $P > 0.5$, $N = 65$).

The strength of the microstimulation effect was significantly correlated with the DTI of the stimulated neurons ($R = 0.44$, $P < 0.001$), with no significant difference between the two monkeys (analysis of covariance, $P = 0.08$). Statistically significant microstimulation effects were generally limited to sites with moderate to strong disparity selectivity. This correlation is reassuring because it provides an additional internal control for nonspecific effects of microstimulation: the effects at weakly tuned sites should be small, and they are.

In the experiments described above, dots always moved in the preferred direction of the stimulated MT neurons. We also tested whether the effects of microstimulation on depth judgements were limited to moving stimuli. We trained monkey R to perform the depth-discrimination task (Fig. 2) when all dots were stationary. Thus, the only image motion on the retina would be that induced by small eye movements made during fixation. We used stationary dots to study 12 MT sites at which microstimulation produced a significant effect when moving dots were used. For 9 of these 12 sites, microstimulation also produced a significant preferred bias when using stationary dots; at the remaining 3 sites there was no significant effect. The mean leftward shift of the psychometric function across all 12 sites was 25.8% for moving dots and 27.9%

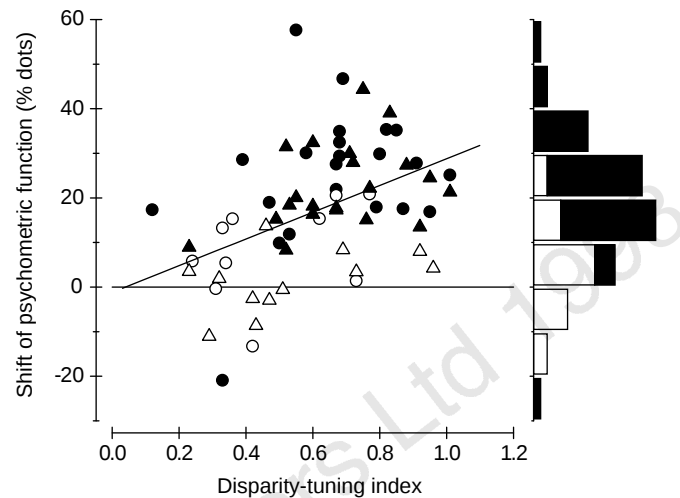


Figure 4 Summary of microstimulation effects from 65 experiments using two monkeys. The scatter plot shows the shift of the psychometric function plotted against the disparity-tuning index of multiunit activity measured at each stimulation site. Filled symbols denote significant effects (logistic regression, $P < 0.05$). Triangles and circles indicate data from monkeys R and T, respectively. The solid line shows the best linear fit to the data (linear regression). On the right, the data are collapsed across the disparity-tuning index and plotted as a histogram.

for stationary dots ($P > 0.5$; paired t -test). We also measured disparity-tuning curves at these 12 sites using stationary dots. All sites exhibited strong disparity selectivity for stationary dots (mean DTI, 0.93 for stationary and 0.70 for moving dots; $P < 0.001$, paired t -test), although the peak response rate was generally smaller for stationary dots (223 events per second) than for moving dots (371 events per second) ($P < 0.001$, paired t -test). Some of the response to stationary dots seemed to be driven by small eye movements during fixation. As these occur under natural viewing conditions, MT neurons may normally be active during viewing of stationary objects.

Our results show that the disparity signals carried by MT neurons are an important part of the neural substrate for depth judgements in our task. This finding establishes the first direct link between the neural property of disparity selectivity and the perceptual capacity for stereopsis, and complements the results of recent studies using both psychophysics and physiology that found a role for MT in the perception of depth from motion cues^{17,18}. Our results also show that the role of MT in vision is not limited to analysis of moving objects, because many MT neurons exhibit robust disparity tuning in response to stationary random-dot patterns, and microstimulation biases depth judgements of stationary dots. Although MT is important in stereopsis under the conditions of our experiment, this does not mean that all aspects of stereopsis depend on MT, nor that any single aspect of stereopsis depends entirely on MT. In this respect, it is interesting to note that a previous study, using a substantially different stereo task, found no effect of MT lesions on psychophysical performance¹⁹. Broadly based investigations, combining a variety of perceptual tests with multiple physiological probes of neural function, will be necessary to achieve a comprehensive understanding of the neural basis of stereoscopic vision. □

Methods

Our methods for physiological recording and microstimulation in awake monkeys have been described previously^{13,20}. We used three rhesus monkeys, two males and one female. During experimental sessions, the monkey was seated in a primate chair with its head restrained, and eye position was monitored using the scleral search coil technique^{21,22}. Two monkeys (R and S) had eye coils implanted in both eyes, to allow monitoring of vergence eye

movements. Systematic changes in vergence angle were not observed as a result of any of the experimental manipulations described below. Tungsten micro-electrodes (impedance typically 250–750 k Ω at 1 kHz) were inserted into the cortex through a transdural guide tube, and event times were stored with 1-ms resolution. Any deflection of the neural signal that exceeded a threshold level was classified as a multiunit event using a bilevel window discriminator. Thus, an increase in amplitude of the neural 'hash' was registered as an increase in frequency of the multiunit activity, which we express in events per second. The absolute frequency of the multiunit response depended on the level of the event threshold and the frequency passband of the amplifier (which was fixed). We attempted to adjust the threshold level to maintain a roughly constant frequency of spontaneous activity from site to site within a penetration. In some experiments (for example, that shown in Fig. 1a), well-isolated action potentials were simultaneously recorded as single units using a template-based spike discriminator. These spikes were excluded from the multiunit response by setting the upper level of the window discriminator below the peak of the spikes. In general, we found the multiunit response to be an excellent predictor of the disparity tuning of single units that were recorded simultaneously at 43 sites in monkey S. Tuning curves constructed from multiunit responses also agreed closely with those computed from the root-mean-square amplitude of the local field potential (bandpass filtered from 20–200 Hz). Monkeys received a liquid reward for correct completion of each trial. Animal care and all experimental procedures conformed to guidelines established by the National Institutes of Health.

Disparity-tuning measurements. The monkey was simply required to maintain fixation on a small spot while a moving random-dot pattern appeared over the multiunit receptive field. Location, size and movement speed of the random-dot pattern were tailored to the preference of the neurons at each recording site. The dots were rendered in depth as red/green anaglyphs viewed at a distance of 57 cm through red and green filters (Kodak Wratten numbers 29 and 61, respectively). Outside the receptive field, the remainder of the visual display (which subtended 30° × 23°) was filled with zero-disparity dots, which were replotted randomly at 20 Hz to produce a flickering background. The zero-disparity background helped to maintain the monkey's vergence at the depth of the fixation point. The video display was refreshed at 60 Hz.

In the experiment shown in Fig. 1b, disparity tuning was measured along an oblique penetration through MT. We approached MT from the occipital lobe; electrodes travelled caudal to rostral in a sagittal plane. The electrode trajectory was tilted 20° away from horizontal. Thus, our electrode passed through MT at an oblique angle, ranging from ~45–90° away from the normal. We recorded multiunit activity at regularly spaced intervals of 100 μ m. When possible, we also isolated spikes from a single unit and recorded these on a separate channel (Fig. 1a).

Disparity-tuning curves were fitted with a cubic spline interpolation. The DTI was defined as: $1 - (R_{\min} - S)/(R_{\max} - S)$, where $R_{\max}$ is the maximum fitted response, $R_{\min}$ is the minimum fitted response, and S denotes spontaneous activity (in the absence of a visual stimulus). Large values of DTI (near 1.0) correspond to strong disparity tuning, and values near zero correspond to weak tuning. For sites with significant disparity tuning (one-way analysis of variance $P < 0.05$), the disparity at which the fitted curve reached a maximum defined the preferred disparity.

Microstimulation experiments. We trained two monkeys (R and T) to perform the depth-discrimination task shown in Fig. 2. To identify candidate microstimulation sites, we searched for regions of MT where the disparity tuning of multiunit activity was roughly constant over at least 200–300 μ m (for example, sites 3–6 in Fig. 1b), and we positioned our electrode in the middle of that region. Microstimulation parameters were similar to those used in previous studies of motion discrimination^{13,23}. The stimulation train consisted of 20- μ A biphasic pulses, each consisting of a 200- μ s cathodal pulse followed by a 200- μ s anodal pulse, with a 100- μ s interval in between. Visual stimuli were presented stereoscopically by alternating left and right half-images at 100 Hz while the monkey viewed the display through ferro-electric shutter glasses that were synchronized to the video refresh. The random-dot stimulus was displayed using the red gun on a standard RGB display, because the red phosphor has the fastest decay time and allows the best stereo separation.

Received 13 April; accepted 1 June 1998.

- Barlow, H. B., Blakemore, C. & Pettigrew, J. D. The neural mechanism of binocular depth discrimination. *J. Physiol. (Lond.)* **193**, 327–342 (1967).
- Pettigrew, J. D., Nikara, T. & Bishop, P. O. Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp. Brain Res.* **6**, 391–410 (1968).
- Hubel, D. H. & Wiesel, T. N. Stereoscopic vision in macaque monkey. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature* **225**, 41–42 (1970).
- Poggio, G. F. & Fischer, B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.* **40**, 1392–1405 (1977).
- Poggio, G. F., Gonzalez, F. & Krause, F. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J. Neurosci.* **8**, 4531–4550 (1988).
- Felleman, D. J. & Van Essen, D. C. Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *J. Neurophysiol.* **57**, 889–920 (1987).
- Maunsell, J. H. & Van Essen, D. C. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* **49**, 1148–1167 (1983).
- Roy, J. P., Komatsu, H. & Wurtz, R. H. Disparity sensitivity of neurons in monkey extrastriate area MST. *J. Neurosci.* **12**, 2478–2492 (1992).
- Cumming, B. G. & Parker, A. J. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* **389**, 280–283 (1997).
- Masson, G. S., Busetini, C. & Miles, F. A. Vergence eye movements in response to binocular disparity without depth perception. *Nature* **389**, 283–286 (1997).
- Albright, T. D. Cortical processing of visual motion. *Rev. Oculomotor Res.* **5**, 177–201 (1993).
- Albright, T. D., Desimone, R. & Gross, C. G. Columnar organization of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.* **51**, 16–31 (1984).
- Salzman, C. D., Murasugi, C. M., Britten, K. H. & Newsome, W. T. Microstimulation in visual area MT: effects on direction discrimination performance. *J. Neurosci.* **12**, 2331–2355 (1992).
- Salzman, C. D. & Newsome, W. T. Neural mechanisms for forming a perceptual decision. *Science* **264**, 231–237 (1994).
- DeAngelis, G. C., Groh, J. M. & Newsome, W. T. Organization of disparity selectivity in macaque area MT. *Soc. Neurosci. Abstr.* **22**, 717 (1996).
- Cox, D. R. & Snell, E. J. *Analysis of Binary Data* (Chapman and Hall, London, 1989).
- Bradley, D. C., Chang, G. C. & Andersen, R. A. Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* **392**, 714–717 (1998).
- Dodd, J. V., Cumming, B. G., Newsome, W. T. & Parker, A. J. Firing of V5 (MT) neurons reliably covaries with reported 3-D configuration in a perceptually-ambiguous structure-from-motion task. *Soc. Neurosci. Abstr.* **23**, 1125 (1997).
- Schiller, P. H. The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Vis. Neurosci.* **10**, 717–746 (1993).
- Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
- Robinson, D. A. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* **10**, 137–145 (1963).
- Judge, S. J., Richmond, B. J. & Chu, F. C. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.* **20**, 535–538 (1980).
- Murasugi, C. M., Salzman, C. D. & Newsome, W. T. Microstimulation in visual area MT: effects of varying pulse amplitude and frequency. *J. Neurosci.* **13**, 1719–1729 (1993).

Acknowledgements. We thank J. Stein and C. Doane for technical assistance, and A. Parker, J. Dodd, B. Wandell, J. Nichols, E. Siedemann, G. Horwitz, and C. Barberini for critical review of the manuscript. G.C.D. was supported by a Medical Research Fellowship from the Bank of America/Giannini Foundation, an NRSA from the National Eye Institute, and a Career Award in the Biomedical Sciences from the Burroughs-Wellcome Fund. B.G.C. is a Royal Society Research Fellow, and W.T.N. is an Investigator of the Howard Hughes Medical Institute. This work was also supported by the National Eye Institute.

Correspondence and requests for materials should be addressed to W.T.N. (e-mail: bill@monkeybiz.stanford.edu).

Maintenance of late-phase LTP is accompanied by PKA-dependent increase in AMPA receptor synthesis

Asha Nayak^{*†}, Devon J. Zastrow^{*†}, Ronald Lickteig[‡], Nancy R. Zahniser^{†‡} & Michael D. Browning^{†‡}

[‡] Department of Pharmacology and [†] Neuroscience Program, University of Colorado Health Sciences Center, 4200 East Ninth Avenue, Denver, Colorado 80262, USA

^{*} These authors contributed equally to this work

Long-term potentiation (LTP) is a form of synaptic plasticity that has been extensively studied as a putative mechanism underlying learning and memory. A late phase of LTP occurring 3–5 hours after stimulation and depending on transcription, protein synthesis and cyclic-AMP-dependent protein kinase (protein kinase A, or PKA) has been described^{1–3}, but it is not known whether transcription of presynaptic and/or postsynaptic genes is required to support late-phase LTP. Here we show that late-phase LTP can be obtained in rat hippocampal CA1 mini-slices

in which the cell bodies of presynaptic Schaffer collateral/commissural fibres are removed. Thus, transcription of presynaptic genes is not necessary to support maintenance of late-phase LTP. The AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionate) receptor is the predominant mediator of the ionotropic response to synaptically released glutamate in the hippocampus and it has been implicated in LTP maintenance. We find that synthesis of AMPA receptor subunits is increased three hours after LTP induction: this effect on the synthesis of the AMPA receptor is blocked by inhibitors of PKA and of transcription. Our results support the idea of a postsynaptic mechanism maintaining late-phase LTP, in which AMPA receptor synthesis is increased as a result of PKA-dependent gene transcription.

To study the biochemistry of LTP, we used a CA1 mini-slice preparation. LTP is produced in these slices using a multi-electrode 'rake' stimulator, which potentiates a much larger fraction of CA1 synapses than is possible with conventional single-electrode stimulation⁴. We first tested whether late-phase LTP could be elicited in this preparation in which the cell bodies of the presynaptic fibres have been removed. As shown in Fig. 1, late-phase LTP can be reliably elicited in the CA1 mini-slice (field excitatory postsynaptic potential (fEPSP) slope was $77 \pm 28\%$ above control at 5 h). This potentiation was fully blocked by the selective PKA inhibitors H-89 (fEPSP slope was $15 \pm 22\%$ above control at 5 h) and RpCAMPS (not shown). Thus, transcription of presynaptic genes is not necessary for late-phase LTP. As late-phase LTP is absent in dendrites excised from CA1 pyramidal somata⁵, transcription of postsynaptic genes may be required to maintain LTP in its later phases: we therefore investigated which postsynaptic proteins might show increased expression in late-phase LTP.

We focused on the synthesis of the AMPA receptor as a likely candidate for a postsynaptic protein whose increased synthesis could be expected to support LTP maintenance during the late phase. Synthesis of AMPA receptor subunits was measured during different phases of LTP by assaying ³⁵S-methionine/cysteine incorporation into the glutamate receptor GluR1 and GluR2/3 subunits at different times after LTP induction. Receptor subunits were individually immunoprecipitated using subunit-specific antibodies against GluR1 and GluR2/3 (ref. 6). As shown in Fig. 2, the synthesis of both GluR1 and GluR2/3 was significantly higher 180 min after stimulation than it was in paired unstimulated controls (GluR1, stimulated: $54 \pm 10\%$ above control, $P < 0.001$; GluR2/3, stimulated: $29 \pm 5\%$ above control, $P < 0.01$, respectively; $n = 10$). These effects were blocked when slices were stimulated in the presence of the NMDA receptor antagonist AP5 (100 μ M), indicat-

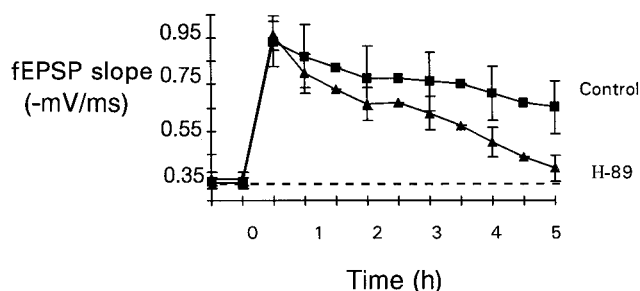


Figure 1 Time course of LTP at various times after induction in CA1 mini-slices incubated in the absence (squares, $n = 6$) or presence (triangles, $n = 7$) of the PKA inhibitor H-89 (10 μ M). Shown are the fEPSP slopes (mean $\pm$ s.e.m.) at various times before and after induction of LTP for slices incubated in the absence or presence of 10 μ M H-89. Two-factor repeated measures ANOVA (treatment vs time) showed a significant interaction ($F_{1,110} = 2.6$, $P < 0.01$). Post-hoc analysis showed that at the 5-h time point the H-89 group was significantly different from control (two-sample t -test, $P < 0.05$).

ing that these effects are indeed LTP-associated. At 90 min, synthesis of both subunits was increased, but neither effect was statistically significant. We next tested whether the LTP-related increase in AMPA receptor synthesis was associated with subcellular fractions enriched in either plasma membranes or microsomes. To this end, we analysed ³⁵S-methionine/cysteine incorporation into GluR1 and GluR2/3 immunoprecipitated from subcellular fractions prepared from control and stimulated CA1 mini-slices. The results show (Fig. 2b) that induction of late-phase LTP was associated with a significant increase in ³⁵S-methionine/cysteine incorporation into GluR1 and GluR2/3 subunits in the subcellular fraction enriched in plasma membranes (GluR1, $55 \pm 11\%$ above control; GluR2/3, $35 \pm 15\%$ above control; $P < 0.01$). No significant LTP-related increase was seen in AMPA receptor synthesis in the fraction enriched in microsomes.

As the synthesis of both GluR1 and GluR2/3 was increased in late-phase LTP, it is possible that late-phase LTP might lead to a nonspecific increase in protein synthesis. We therefore investigated the synthesis of three additional proteins, the NMDA receptor subunits NR2A and NR2B and the presynaptic protein synapsin I. The synthesis of synapsin I and the NMDA receptor subunits was

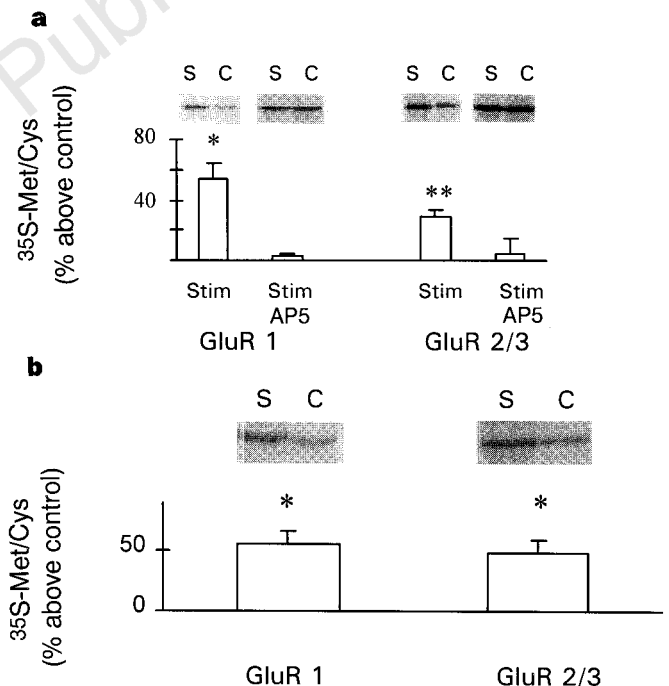


Figure 2 AMPA receptor synthesis is increased in late-phase LTP. **a**, ³⁵S-methionine/cysteine incorporation into immunoprecipitated GluR1 and GluR2/3 subunits of the AMPA receptor is significantly increased 180 min after LTP induction. Two-factor repeated measures ANOVA (treatment vs time) for each subunit showed a significant effect of stimulation for both GluR1 ($F_{1,26} = 13.93$, $P < 0.001$) and GluR2/3 ($F_{1,27} = 17.10$, $P < 0.001$). Post-hoc analysis showed that synthesis of both GluR1 and GluR2/3 was significantly increased 180 min after stimulation (paired t -test, $*P < 0.001$, $**P < 0.01$ respectively). Autoradiograms of a representative stimulated (S) and control (C) slice pairs are shown for each condition: band intensity reflects the relative ³⁵S-methionine/cysteine incorporation into individual subunits. Apparent differences in autoradiograph band intensities between controls of different groups are due only to differences in film exposure time. **b**, Synthesis of AMPA receptor subunits GluR1 and GluR2/3 is significantly higher in plasma-membrane-enriched fractions prepared from stimulated slices (paired t -test, $*P < 0.01$). After preparation of plasma-membrane- and microsomal-enriched fractions, ³⁵S-methionine/cysteine incorporation into immunoprecipitated GluR1 and GluR2/3 subunits of the AMPA receptor was determined. No significant increase was seen in AMPA receptor subunit synthesis in the fraction enriched in microsomes.

not changed during late-phase LTP (Fig. 3), suggesting that the increase in synthesis in late-phase LTP is relatively specific for AMPA receptor subunits. To confirm specificity of these effects, we analysed protein bands with relative molecular masses (M_r) of 200K, 60K, 52K, 46K and 38K seen on SDS-PAGE of total homogenate. There was no stimulation-dependent increase in ^{35}S -methionine incorporation into any of these protein bands (not shown). To verify that LTP-inducing stimulation was required to increase AMPA receptor synthesis, we delivered the same number of stimuli at the same positions in the slice, but lowered the stimulation frequency from 100 Hz to 3 Hz, a frequency that does not produce LTP. Such stimulation had no effect on the synthesis of the AMPA receptor subunits (^{35}S -methionine incorporation for GluR1 was $-7.5 \pm 16\%$ above control, and for GluR2/3 was $6.1 \pm 21\%$ of control; means $\pm$ s.e.m., $n = 8$).

To eliminate the possibility that the increase in ^{35}S -methionine/cysteine incorporation merely reflects an increase in receptor turnover, we compared total levels of AMPA receptor subunits in stimulated and unstimulated control slices using quantitative western blot analysis. As shown in Fig. 4, GluR1 and GluR2/3 levels were significantly increased 180 min after LTP induction (GluR1 levels, $16 \pm 7\%$ above control, $P < 0.02$; GluR2/3 levels, $13 \pm 6\%$ above control, $P < 0.05$). To test whether there was an increase in AMPA receptor binding during late-phase LTP, we did a saturation analysis of ^3H -AMPA binding⁷ in crude particulate fractions prepared from control and stimulated mini-slices. Our results show that at 180 min after LTP induction there was an average increase in ^3H -AMPA B_{max} of 17.6% in stimulated slices, a value that compares favourably with the increase in the amount of receptor determined from western-blot analysis. However, there was substantial variability in this assay owing to the limited tissue available in the mini-slice and the increase in B_{max} was not significantly significant (B_{max} in stimulated slices was

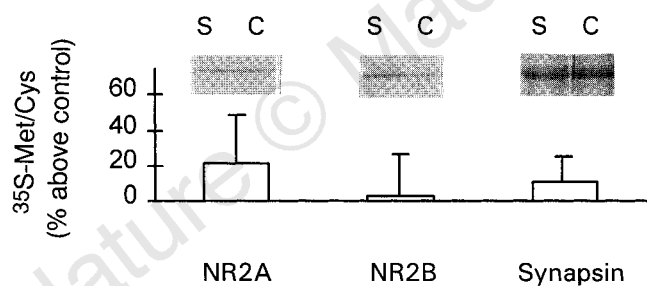


Figure 3 ^{35}S -methionine/cysteine incorporation into immunoprecipitated NMDA receptor subunits NR2A, NR2B or synapsin is not significantly increased 180 min after LTP induction. Bar graphs show cumulative data expressed as the mean percentage increase over control $\pm$ s.e.m. in ^{35}S -methionine/cysteine incorporation into immunoprecipitated proteins.

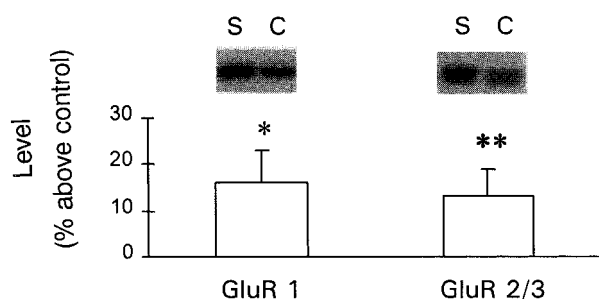


Figure 4 Quantitative western blot analysis reveals an increase in AMPA receptor subunits 180 min after LTP induction. Bar graphs of cumulative data show the mean percentage increase $\pm$ s.e.m. in subunit expression in stimulated slices above control levels 180 min following stimulation. A significant increase in GluR1 ($16 \pm 7\%$, $P < 0.02$; $n = 13$) and GluR2/3s ($13 \pm 6\%$, $P < 0.05$; $n = 13$) was seen.

$515 \pm 72 \text{ fmol mg}^{-1}$ protein, and in control slices it was $438 \pm 52 \text{ fmol mg}^{-1}$ protein; means $\pm$ s.e.m.; $n = 6$).

As late-phase LTP is dependent on PKA activation², we tested whether increased AMPA receptor synthesis might also depend on such activation by assaying for receptor synthesis in mini-slices stimulated in the presence of either RpcAMPS or H-89: we found that the stimulation-dependent increases in the synthesis of both GluR1 and GluR2/3 were blocked (Fig. 5a). Last, because late-phase LTP is blocked by inhibitors of gene transcription³, we determined whether the observed increase in AMPA receptor synthesis required the transcription of new messenger RNAs. As shown in Fig. 5b, the increase was completely abolished in slices stimulated in the presence of the transcriptional inhibitors actinomycin D or 5,6-dichlorobenzimidazole riboside (DRB).

Although the late phase of LTP is known to depend on transcription, protein synthesis and PKA activation, little is known about whether pre- or postsynaptic genes are involved or about which protein(s) is(are) synthesized to support this stage of LTP. We have shown here that presynaptic gene transcription is not necessary for late-phase LTP. Thus, late-phase LTP in the rat hippocampus may differ from forms of long-lasting plasticity in *Aplysia* and *Drosophila*, in which transcription of presynaptic genes appears to be necessary⁸⁻¹⁰. Moreover, we find that during late-phase LTP there is a relatively specific increase in the synthesis of AMPA receptor subunits that is blocked by inhibitors of either PKA or of transcription. Our results are consistent with the idea that late-phase LTP depends, at least in part, on an increase in postsynaptic expression of AMPA receptors.

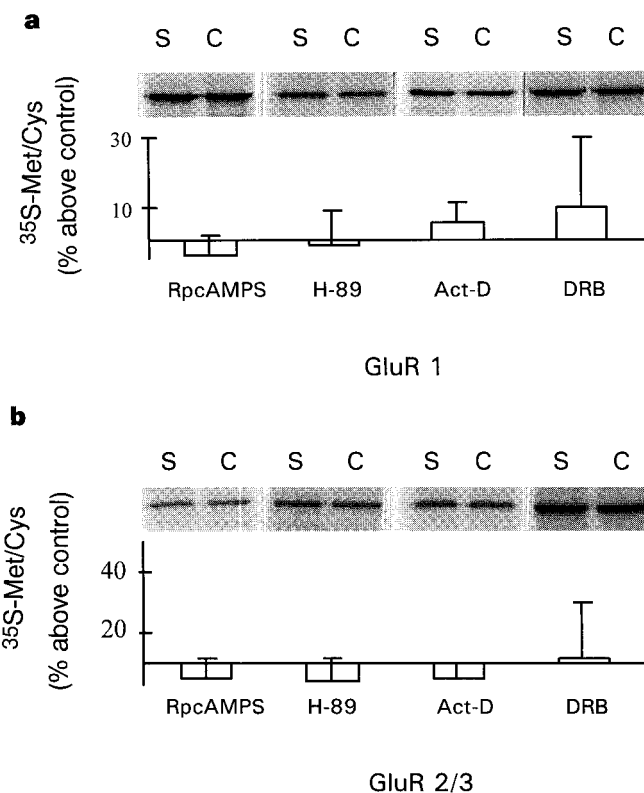


Figure 5 LTP-associated increased in GluR synthesis are fully blocked by RpcAMPS, H-89, actinomycin D and DRB. ^{35}S -methionine/cysteine incorporation into newly synthesized GluR1 (a) and GluR2/3 (b) was measured 180 min after rake-LTP stimulation and expressed as a percentage above unstimulated control levels. Stimulation-dependent increases in synthesis were blocked by $100 \mu\text{M}$ RpcAMPS (GluR1, $-4 \pm 6\%$; GluR2/3, $-5 \pm 7\%$, $n = 12$), $10 \mu\text{M}$ H-89 (GluR1, $-1 \pm 10\%$; GluR2/3, $-6 \pm 8\%$, $n = 8$), 25 mM actinomycin D (GluR1, $5 \pm 6\%$, GluR2/3, $-5 \pm 5\%$, $n = 12$) and $100 \mu\text{M}$ DRB (GluR1, $-10 \pm 20\%$; GluR2/3 $2 \pm 18\%$, $n = 8$). Values are mean $\pm$ s.e.m.

The mechanism by which newly synthesized AMPA receptors might support the maintenance of LTP in its late phase is not known. Our subcellular fractionation data indicate that newly synthesized AMPA receptors may be incorporated into synaptic membranes to support late-phase LTP. These newly synthesized receptors could insert into pre-existing functional AMPA synapses and augment responses at those synapses. Alternatively, newly synthesized receptors could insert into 'silent synapses' and result in functional AMPA receptor responses at such synapses^{11–13}. These mechanisms might replace the phosphorylation of the AMPA receptor that occurs in early-phase LTP¹⁴ as a potentiating mechanism in late-phase LTP. Such models remain speculative and are a basis for future investigation of late-phase LTP. □

Methods

Hippocampal mini-slice preparation. Hippocampal CA1 mini-slices were prepared as described¹⁵ and perfused with artificial-cerebrospinal-fluid buffer (ACSF: 124 mM NaCl, 4 mM KCl, 1 mM MgCl₂, 2.5 mM CaCl₂, 1 mM KH₂PO₄, 10 mM dextrose, 25.7 mM NaHCO₃). The rake electrode was prepared as a linear array of four 50 µm diameter, Teflon-insulated, stainless-steel monopolar electrodes, each separated by 100 µm and driven by its own stimulus isolation unit. Rake-LTP was induced by the delivery of 3 stimulus trains (100 Hz for 1 s, 30 s apart). For all LTP stimulation experiments, three trains were delivered at 4 separate positions ~0.75 mm apart within the CA1 mini-slice. Immediately following stimulation, stimulated and control slices were labelled in 1 mCi ml⁻¹ ³⁵S-methionine/cysteine for 90 or 180 min. All drugs were applied for 20 min before stimulation and throughout the labelling or incubation periods.

Immunoprecipitation. Slices were sonicated in 1% SDS and subjected to SDS-PAGE, or specific proteins were immunoprecipitated as described¹⁶ using antibodies specific for GluR1 or GluR2/3, NMDA R2A, NMDA R2B or synapsin. ³⁵S-incorporation into specific subunits was quantified using a BioRad PhosphorImager. The specificity of the GluR (Chemicon), NMDA and synapsin antibodies has been described^{6,17,18}. For all immunoprecipitation figures, autoradiograms of a representative stimulated (S) and control (C) slice pair are shown for each condition, where band intensity reflects the relative level of ³⁵S-methionine/cysteine incorporation into individual subunits. Bar graphs show cumulative data expressed as the mean percentage increase over control ± s.e.m. in ³⁵S-methionine/cysteine incorporation into immunoprecipitated proteins.

Western blots. Quantitative western blots were prepared as previously described¹⁷. Blots containing equal amounts of total protein from stimulated and control slices were probed for GluR1 and GluR2/3 levels using quantitative western analysis conditions established for each antibody: 0.25 mg ml⁻¹ GluR1 antibody and 0.10 mg ml⁻¹ GluR2/3. Immune complexes were detected using ¹²⁵I-labelled protein A (100 nCi ml⁻¹). ¹²⁵I-protein A labelling of GluR subunits was linear with protein load in this analysis.

Subcellular fractionation. After incubation in ³⁵S-methionine/cysteine mini-slices were pooled in groups of three stimulated and control matched pairs. The slices were washed twice in 4 ml fresh, oxygenated ACSF, then again in 100 µl homogenization buffer (10 mM Tris, pH 7.5, 10 mM EGTA, 10 mM EDTA, 0.1% (v/v) β-mercaptoethanol (BME), 0.1% (v/v) aprotinin, 4 mg ml⁻¹ leupeptin, 4 mg ml⁻¹ soybean trypsin inhibitor (SBTI), 4 mg ml⁻¹). Slices were homogenized in 150 µl homogenization buffer and allowed to lyse on ice for 30 min. Fractions were spun for 10 min at 900g. The supernatants were removed and spun 15 min at 12,000g; the resulting supernatants were removed to another tube and the pellet was sonicated in 50 µl additional lysis buffer. This fraction was again spun 15 min at 12,000g. Supernatants were then pooled (microsomal enriched fraction) and 20 µl 10% SDS was added. Pellets (plasma-membrane-enriched fraction) were sonicated in 50 µl 1% SDS.

Received 18 May; accepted 10 June 1998.

1. Frey, U., Krug, M., Reymann, K. G. & Matthies, H. Anisomycin, an inhibitor of protein synthesis, blocks late phases of LTP phenomena in the hippocampal CA1 region *in vitro*. *Brain Res.* **452**, 57–65 (1988).
2. Frey, U., Huang, Y.-Y. & Kandel, E. R. Effects of cAMP simulate a late stage of LTP in hippocampal CA1 neurons. *Science* **260**, 1661–1664 (1993).
3. Nguyen, P. V., Abel, T. & Kandel, E. R. Requirement of a critical period of transcription for induction of a late-phase of LTP. *Science* **265**, 1104–1107 (1994).

4. Nayak, A. S., Moore, C. I. & Browning, M. D. CAM kinase II phosphorylation of the presynaptic protein synapsin is persistently increased during expression of long-term potentiation. *Proc. Natl Acad. Sci. USA* **93**, 15451–15456 (1996).
5. Frey, U., Krug, M., Brodemann, R., Reymann, K. & Matthies, H. Long-term potentiation induced in dendrites separated from rat's CA1 pyramidal somata does not establish a late phase. *Neurosci. Lett.* **97**, 135–139 (1989).
6. Wenthold, R. J., Yokotani, N., Doi, K. & Wada, K. Immunohistochemical characterization of the non-NMDA glutamate receptor using subunit-specific antibodies. Evidence for a hetero-oligomeric structure in rat brain. *J. Biol. Chem.* **267**, 501–507 (1992).
7. Maren, S., Tocco, G., Standley, S., Baudry, M. & Thompson, R. F. Postsynaptic factors in the expression of long-term potentiation (LTP): Increased glutamate receptor binding following LTP induction *in vivo*. *Proc. Natl Acad. Sci. USA* **90**, 9654–9658 (1993).
8. Bartsch, D. et al. Aplysia CREB2 represses long-term facilitation: Relief of repression converts transient facilitation into long-term functional and structural change. *Cell* **83**, 979–992 (1995).
9. Martin, K. C. & Kandel, E. R. Cell adhesion molecules, CREB, and the formation of new synaptic connections. *Neuron* **17**, 567–570 (1996).
10. Davis, G. W., Schuster, C. M. & Goodman, C. S. Genetic dissection of structural and functional components of synaptic plasticity. 3. CREB is necessary for presynaptic functional plasticity. *Neuron* **17**, 669–679 (1996).
11. Liao, D., Hessler, N. A. & Malinow, R. Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. *Nature* **375**, 400–404 (1995).
12. Voronin, L. L., Volgushev, M., Chistiakova, M., Kuhnt, U. & Singer, W. Involvement of silent synapses in the induction of long-term potentiation and long-term depression in neocortical and hippocampal neurons. *Neuroscience* **74**, 323–330 (1996).
13. Isaac, J. T. R., Crair, M. C., Nicoll, R. A. & Malenka, R. C. Silent synapses during development of thalamocortical inputs. *Neuron* **18**, 269–280 (1997).
14. Barria, A., Muller, D., Derkach, V., Griffith, L. C. & Soderling, T. R. Regulatory phosphorylation of AMPA-type glutamate receptors by CaM-KII during long-term potentiation. *Science* **276**, 2042–2045 (1997).
15. Parfitt, K. D., Doze, V. A., Madison, D. V. & Browning, M. D. Isoproterenol increases the phosphorylation of the synapsins and increases synaptic transmission in dentate gyrus but not area CA1 of the hippocampus. *Hippocampus* **2**, 59–64 (1992).
16. Haycock, J. W., Browning, M. D. & Greengard, P. Cholinergic regulation of protein phosphorylation in bovine adrenal chromaffin cells. *Proc. Natl Acad. Sci. USA* **85**, 1677–1681 (1988).
17. Stone, L. M., Browning, M. D. & Finger, T. E. Differential distribution of the synapsins in the rat olfactory bulb. *J. Neurosci.* **14**, 301–309 (1994).
18. Snell, L. D. et al. Regional and subunit specific changes in NMDA receptor mRNA and immunoreactivity in mouse brain following chronic ethanol ingestion. *Mol. Brain Res.* **40**, 71–78 (1996).

Calcium-permeable AMPA receptors mediate long-term potentiation in interneurons in the amygdala

Nishith K. Mahanty & Pankaj Sah*

The Neuroscience Group and the Discipline of Human Physiology, Faculty of Medicine and Health Sciences, University of Newcastle, NSW, 2308, Australia

Fear conditioning is a paradigm that has been used as a model for emotional learning in animals¹. The cellular correlate of fear conditioning is thought to be associative *N*-methyl-D-aspartate (NMDA) receptor-dependent synaptic plasticity within the amygdala^{1–3}. Here we show that glutamatergic synaptic transmission to inhibitory interneurons in the basolateral amygdala is mediated solely by α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors. In contrast to AMPA receptors at inputs to pyramidal neurons, these receptors have an inwardly rectifying current–voltage relationship, indicative of a high permeability to calcium^{4,5}. Tetanic stimulation of inputs to interneurons caused an immediate and sustained increase in the efficacy of these synapses. This potentiation required a rise in postsynaptic calcium, but was independent of NMDA receptor activation. The potentiation of excitatory inputs to interneurons was reflected as an increase in the amplitude of the GABA_A-mediated inhibitory synaptic current in pyramidal neurons. These results demonstrate that excitatory synapses onto interneurons within a fear conditioning circuit show NMDA-receptor independent long-term potentiation. This plasticity might underlie the increased synchronization of activity between neurons in the basolateral amygdala after fear conditioning⁶.

* Present address: Division of Neuroscience, John Curtin School of Medical Research, GPO Box 334, Canberra ACT 2601, Australia.

Whole-cell recordings were made from neurons in the lateral amygdala and basolateral amygdala. Two classes of cells were encountered on the basis of their electrophysiological properties^{7,8}. One class had broad action potentials with half-width 1.2 ± 0.1 ms ($n = 50$, Fig. 1a) and, in response to a prolonged current injection, fired trains of action potentials that showed varying degrees of spike frequency adaptation (Fig. 1c). A large slow after-hyperpolarizing potential (AHP) always followed trains of action potentials (not shown). These cells were classified as pyramidal cells. In contrast, a second cell type had faster action potentials with half-width 0.76 ± 0.04 ms ($n = 8$; Fig. 1b) and no appreciable spike frequency adaptation (Fig. 1d). No slow AHPs were apparent in these cells. These cells were classified as interneurons.

In voltage clamp, spontaneous inward synaptic currents (sEPSCs) were readily apparent in both pyramidal cells and interneurons (Fig. 1e, f). The amplitude of sEPSCs was larger in interneurons (32.4 ± 2.2 pA, $n = 21$) than in pyramidal cells (15.5 ± 0.8 pA, $n = 17$). sEPSCs were always significantly faster ($P < 0.001$) in interneurons. The 20–80% rise times of sEPSCs were 0.54 ± 0.06 ms ($n = 11$) for interneurons and 1.11 ± 0.05 ms ($n = 24$) for pyramidal neurons. The decay time constants were 2.4 ± 0.2 ms and 7.0 ± 0.4 ms for interneurons and pyramidal cells, respectively (Fig. 1g, h). sEPSCs were abolished on application of the AMPA/kainate receptor antagonist CNQX in both pyramidal cells ($n = 10$; Fig. 2a) and interneurons ($n = 4$; Fig. 2c), showing that they were glutamatergic. Thus, as in other brain regions,

pyramidal cells and interneurons in the lateral amygdala/basolateral amygdala can be reliably identified on the basis of their repetitive firing properties^{9,10} or the kinetics of their glutamatergic sEPSCs¹¹. The majority of neurons (552/596; 93%) were classified as pyramidal neurons, whereas interneurons were far less frequent (44/596; 7%).

To examine the voltage dependence of sEPSCs in the two cell types we recorded from cells by using caesium-based internal solutions to reduce outward potassium currents. In pyramidal neurons, inward sEPSCs were present at -60 mV and outward sEPSCs could be detected at $+60$ mV (Fig. 2b). The fast outward currents were also abolished by 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX), leaving in some cells a slower outward current (not shown). These data are consistent with those in a previous report showing dual component glutamatergic spontaneous synaptic events in pyramidal cells in the basolateral amygdala¹². In contrast to pyramidal cells, no clear outward currents could be recorded in interneurons at $+60$ mV despite similar levels of baseline noise (Fig. 2d). These results suggest that AMPA-mediated synaptic currents in interneurons rectify in the inward direction.

To study this effect more systematically, we examined the voltage dependence of evoked synaptic currents. In both cell types, stimulation of the external capsule evoked fast inward synaptic currents at negative holding potentials (Fig. 3a, d). Upon depolarization of pyramidal neurons a slower current was also recorded that was selectively blocked by the NMDA receptor antagonist 2-amino-5-phosphonopentanoate (D-APV) (Fig. 3c). These two components of the evoked current are typical AMPA- and NMDA-receptor-mediated synaptic currents¹³. The AMPA-receptor-mediated com-

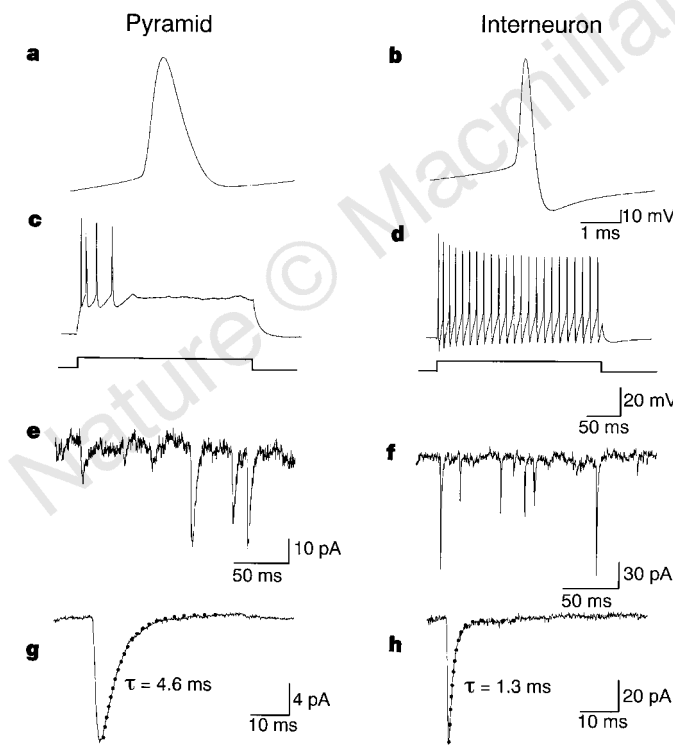


Figure 1 Properties of pyramidal cells and interneurons in the basolateral amygdala. **a, b**, Examples of action potentials recorded from a pyramidal cell (resting membrane potential -58 mV) and interneuron (resting membrane potential -55 mV) respectively in response to depolarizing current injection (0.5 and 0.1 nA for pyramidal and interneuron, respectively). **c, d**, Injection of a prolonged depolarizing current produces repetitive action potentials that show marked spike frequency adaptation in pyramidal cells. In contrast, interneurons do not show any spike frequency adaptation. Current injections were 0.4 and 0.3 nA for pyramidal cell and interneuron respectively. **e, f**, Spontaneous synaptic currents recorded in a pyramidal cell voltage clamped at -60 mV. **g, h**, Average sEPSC in a pyramidal cell and interneuron. The dotted lines are the best-fit single exponential with the indicated time constants.

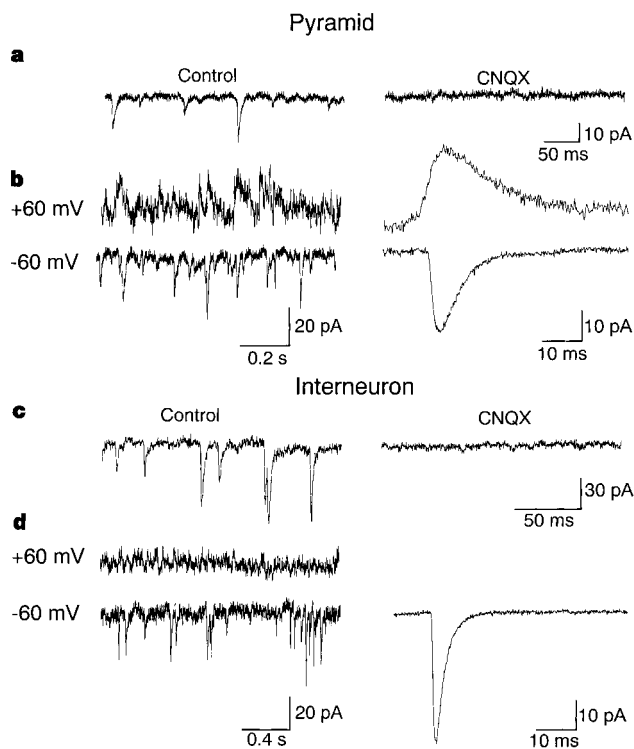


Figure 2 Spontaneous EPSCs in pyramidal cells and interneurons are glutamatergic. **a**, Spontaneous sEPSCs recorded in a pyramidal cell at a holding potential of -60 mV are abolished by the application of $10 \mu\text{M}$ CNQX. **b**, sEPSCs recorded in a different pyramidal cell recorded at holding potentials of -60 mV and $+60$ mV. Average sEPSCs at the two potentials are shown on the right. **c**, sEPSCs recorded in an interneuron at a holding potential of -60 mV are abolished by the application of $10 \mu\text{M}$ CNQX. **d**, sEPSCs in interneurons can be recorded only at a holding potential of -60 mV. No sEPSCs can be recorded at $+60$ mV. The average sEPSCs at -60 mV are shown on the right.

ponent at these synapses had a linear current–voltage (I – V) relation (Fig. 3b). In contrast, depolarization of interneurons did not reveal any slow current (Fig. 3d) and the synaptic current was entirely blocked by CNQX (Fig. 3d, inset). The ratio of the peak AMPA-mediated synaptic current (measured at -60 mV) to that of the NMDA-receptor component (measured 80 ms after the peak at $+40$ mV) was 16 ± 4 ($n = 39$) for pyramidal cells. This ratio was always >1000 for interneurons. These results indicate that the contribution of NMDA receptors at glutamatergic synapses onto interneurons is very small or negligible. This result is in contrast to glutamatergic synapses on interneurons in the hippocampus, where NMDA-receptor components can be readily detected^{14,15} and may reflect the relative paucity of NMDA-receptors in dendritic shafts of neurons in the basolateral amygdala¹⁶.

The peak I – V relation of the AMPA component in interneurons showed a distinct inward rectification (Fig. 3e). The rectification index (RI), defined as the conductance of the AMPA component

measured at $+40$ mV divided by the conductance at -60 mV, varied between 0.5 and 1.7 for pyramidal cells, with an average of 0.90 ± 0.07 ($n = 27$). In comparison, the RI for the AMPA component in interneurons varied between 0.01 and 0.3 with an average of 0.17 ± 0.04 ($n = 13$, $P < 0.0001$; Fig. 3f).

AMPA receptors are heteromultimers assembled from four different subunits¹⁷. The biophysical properties of these receptors are dominated by those of the GluR2 subunit: receptors lacking GluR2 show marked inward rectification and high calcium permeability, whereas receptors assembled with GluR2 have linear or outwardly rectifying I – V values and low calcium permeability^{4,5,18,19}. Glutamatergic synaptic currents in a subset of hippocampal interneurons that have low levels of GluR2 (ref. 20) also show inward rectification^{21,22}. Our data showing an inwardly rectifying I – V for the AMPA-mediated synaptic current in interneurons in the basolateral amygdala are consistent with the low levels of GluR2 expression found in these cells²³. Receptors containing low levels of GluR2 can also be distinguished by their selective block by external polyamines⁴. To assess this we applied NHPP-spermine ($5 \mu\text{M}$; Fig. 3g) while stimulating afferents at 0.1 Hz. At a holding potential of -60 mV, the EPSC in interneurons was blocked by $63 \pm 9\%$ ($n = 3$), whereas the EPSC in pyramidal cells was blocked by only $18 \pm 5\%$ ($n = 5$; $P < 0.001$), consistent with these interneurons' expressing low levels of GluR2.

As AMPA receptors that are blocked by external polyamines and have rectifying I – V relations are associated with a high calcium permeability^{4,19}, we next asked whether calcium influx via these receptors has a functional role. Increases in postsynaptic calcium underlie a number of processes including synaptic plasticity²⁴. We therefore tested for the presence of long-term potentiation (LTP) at these rectifying synapses. After obtaining a stable baseline, inputs to interneurons were tetanized (30 Hz, 100 stimuli, repeated twice) with the cells voltage clamped at -70 mV to minimize the contribu-

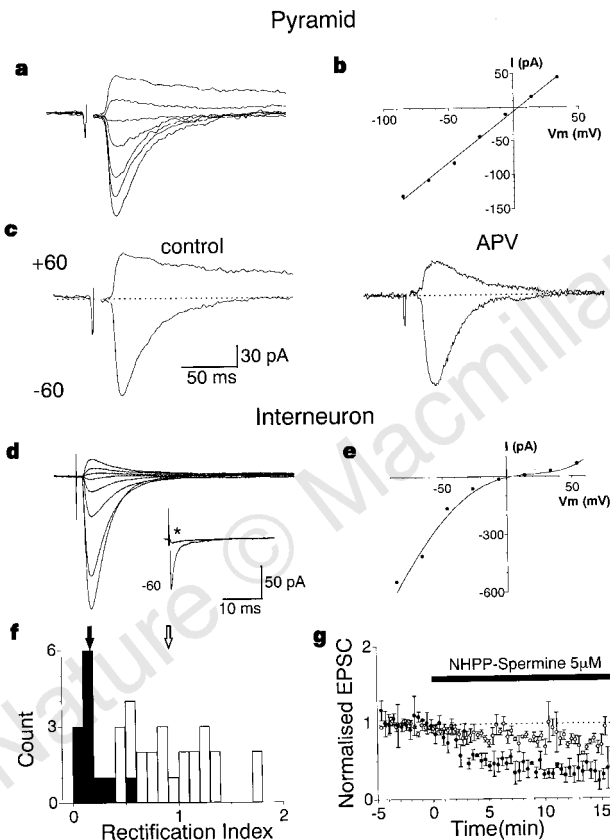


Figure 3 Synaptic currents mediated by AMPA receptors in interneurons are inwardly rectifying. **a**, Evoked synaptic currents recorded from a pyramidal cell at holding potentials of -80 to $+40$ mV. Note the slower component that is revealed with membrane depolarization. **b**, The peak I – V relationship at these synapses is linear. **c**, The NMDA-receptor antagonist d-APV ($30 \mu\text{M}$) blocks the slow component of the synaptic current. **d**, Synaptic currents recorded in an interneuron at holding potentials of -80 mV to $+60$ mV. By contrast with pyramidal cells, evoked synaptic currents in interneurons do not show any NMDA-receptor-mediated component. The inset shows that the synaptic current is entirely blocked by the AMPA receptor antagonist CNQX ($10 \mu\text{M}$, trace marked *). **e**, The peak I – V relation for the AMPA component in interneurons is inwardly rectifying. **f**, RI is plotted for 13 interneurons (filled bars) and 27 pyramidal cells (open bars). The average RI in interneurons was 0.17 ± 0.04 (closed arrow) and 0.90 ± 0.07 in pyramidal cells (open arrow). **g**, Average normalized EPSC amplitudes are plotted from interneurons ($n = 3$; filled circles) and pyramidal cells ($n = 5$; open circles). After obtaining a stable baseline at a holding potential of -60 mV, NHPP-spermine was applied at $5 \mu\text{M}$. AMPA synapses at interneurons were blocked by $63 \pm 9\%$, whereas those in pyramidal cells were blocked by $18 \pm 5\%$.

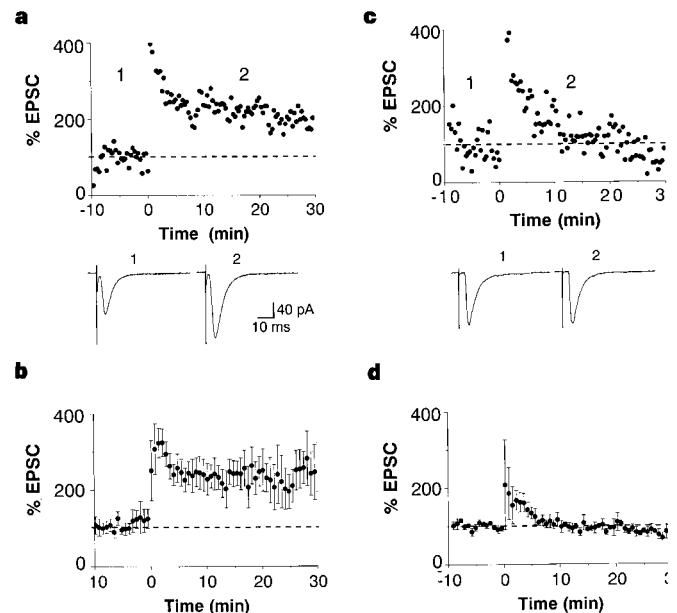


Figure 4 Long-term potentiation in interneurons requires an increase in postsynaptic calcium concentration. **a**, **b**, LTP at interneurons in the amygdala. **a**, EPSC amplitudes plotted for a single cell against time. The afferent fibres were tetanized at time zero. Sweeps recorded at the times indicated are shown below. The average LTP recorded from seven cells is shown in **b**. **c**, **d**, LTP in interneurons requires an increase in postsynaptic calcium concentration. **c**, EPSC amplitudes plotted for a single cell plotted against time for a cell loaded with 10 mM BAPTA. The afferent fibres were tetanized at time zero. Sweeps recorded at the times indicated are shown below. Average data for six cells loaded with BAPTA are shown in **d**.

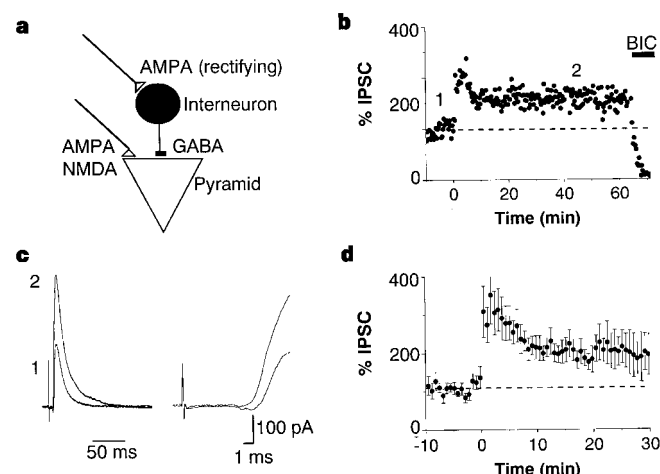


Figure 5 Disynaptic LTP of inhibitory synaptic currents in pyramidal cells. **a**, Diagram showing the circuitry within the basolateral amygdala. Recordings were made from pyramidal neurons. **b**, Peak IPSC amplitude recorded at a membrane potential of -30 mV is plotted against time for a single cell. The afferent fibres were tetanized at time zero. The GABA_A receptor antagonist bicuculline (BIC) ($10 \mu\text{M}$), applied during the time indicated by the solid bar, abolished the IPSC. **c**, Sweep data from the times indicated in **b** are superimposed at slow (left) and fast (right) time bases. LTP of the GABAergic synaptic current is associated with a clear decrease in the latency of the IPSC. **d**, Average data ($n = 7$) for LTP of the inhibitory synaptic current.

tion from voltage-gated calcium channels. After the tetanus, synaptic inputs displayed LTP in 7/7 cells. The average potentiation 30 min after the tetanus was $258 \pm 55\%$ (Fig. 4a, b). Consistent with the lack of NMDA receptors at these synapses, LTP was not blocked in the presence of the NMDA-receptor antagonist D-APV ($50 \mu\text{M}$; $n = 2$). A similar tetanus delivered to pyramidal cells showed no consistent effect ($n = 42$; not shown). To test whether a rise in postsynaptic calcium is necessary to elicit LTP, we loaded cells with the calcium chelator 1,2-bis(*o*-aminophenoxy)-ethane-*N,N,N',N'*-tetraacetic acid (BAPTA, 10 mM) for at least 10 min. LTP was completely blocked in cells loaded with BAPTA ($n = 6$; Fig. 4c, d) showing that an increase in postsynaptic calcium is essential for LTP induction. Pairing synaptic stimulation (30–50 stimuli at 1 Hz) with either depolarization of interneurons to 0 mV ($n = 5$) or with voltage steps (-70 mV to $+10$ mV, 300 ms; $n = 4$) to activate voltage-gated calcium channels had no effect on EPSC amplitude, showing that calcium influx via voltage-gated calcium channels is insufficient to induce LTP.

Interneurons in the amygdala are GABAergic, and stimulation in the external capsule elicits a disynaptic inhibitory postsynaptic potential (Fig. 5a) in pyramidal neurons that is mediated, in part, by GABA_A receptors²⁵. If repetitive stimulation potentiates the EPSC onto these interneurons, then the disynaptic inhibitory synaptic current (IPSC) recorded from pyramidal cells should also be potentiated. This was found to be so (Fig. 5). Disynaptic IPSCs were recorded from pyramidal cells at a holding potential of -30 mV. After a stable baseline had been obtained, a tetanus was delivered to the external capsule. This led to a persistent increase ($215 \pm 65\%$; $n = 7$) in the amplitude of the IPSC (Fig. 5b, d). The increase in IPSC amplitude was associated with a clear decrease in the latency of the IPSC, which is consistent with the potentiation of the IPSC being polysynaptic (Fig. 5c). When monosynaptic IPSCs were recorded in the presence of $10 \mu\text{M}$ CNQX, a similar tetanus failed to potentiate the IPSC, indicating that GABAergic synapses do not exhibit LTP under these conditions ($n = 4$; not shown).

These results show that pyramidal neurons and GABAergic interneurons in the basolateral amygdala receive glutamatergic inputs with distinct properties. Inputs onto pyramidal neurons

activate synapses at which both AMPA and NMDA receptors are active. The AMPA component of the synaptic current at these synapses has a linear I - V , indicating that the underlying receptors contain GluR2 subunits⁴. In contrast, glutamatergic inputs onto interneurons activate AMPA-only synapses that have an inwardly rectifying I - V and are blocked by extracellular NHPP-spermine, indicating that these receptors contain relatively low levels of GluR2 and are calcium-permeable^{4,18}. Tetanic stimulation of synapses onto interneurons elicits LTP, which requires a rise in postsynaptic calcium. This increase in calcium is likely to be due to influx via AMPA receptors^{26,27}. The downstream consequence of this potentiation is an increase in the disynaptic GABAergic IPSC in pyramidal neurons.

The amygdala is a critical component of the circuit that mediates fear conditioning^{1,2}. Associative synaptic plasticity at excitatory inputs to the basolateral amygdala has been proposed to be the cellular mechanism underlying this type of learning, and studies *in vivo* have corroborated this idea^{3,6,28}. We have shown that, after tetanic stimulation, inhibitory inputs onto pyramidal cells are also potentiated. The axons of interneurons diverge widely and thus changes in excitatory drive to interneurons will increase inhibitory potentials in many pyramidal cells. This increase in inhibitory transmission might underlie²⁹ the increased synchronization of activity between neurons in the basolateral amygdala after fear conditioning⁶. □

Methods

Slice preparation. Experiments were performed on coronal brain slices prepared by standard techniques. Rats (Wistar, 17–25 day old), were anaesthetized with intraperitoneal pentobarbitone (50 mg kg^{-1}) and decapitated. Oxygenated, heated (28 – 30°C) Ringer containing (mM) NaCl 118, KCl 2.5, NaHCO_3 25, glucose 10, NaH_2PO_4 1.2, MgCl_2 1.3 and CaCl_2 2.5 was superfused over the slice at a constant rate of 200 ml h^{-1} in a bath volume of 0.5 ml. Picrotoxin ($50 \mu\text{M}$) was included in the perfusing Ringer to block GABA_A-mediated synaptic events when excitatory synaptic transmission was studied.

Whole-cell recording. Whole-cell patch clamp recordings were obtained from neurons in the lateral amygdala and basolateral amygdala with the 'blind' approach³⁰. Electrodes (2 – $4 \text{ M}\Omega$) were filled with a solution containing (in mM) caesium gluconate 135, CsCl 10, NaCl 8, HEPES 10, Mg_2ATP 2, Na_3GTP 0.2, EGTA 0.2 and spermine 0.1 (made to pH 7.3 with CsOH). For current clamp recordings, electrodes contained (in mM) KMeSO_4 135, NaCl 8, HEPES 10, Mg_2ATP 2 and Na_3GTP 0.3 (made to pH 7.3 with KOH, osmolarity 290 mOsm). All membrane potentials have been corrected for a junction potential of $+13 \text{ mV}$ recorded with the caesium gluconate internal.

Stimulation and recording. Stainless-steel bipolar stimulating electrodes (Frederick Haer) were placed on the external capsule. Stimuli were 10 – 30 V in amplitude and $100 \mu\text{s}$ in duration. Signals were recorded with an Axopatch 1D (Axon Instruments), filtered at 5 kHz and digitized at 10 kHz (Instrutech, ITC-16) with custom software running on a Macintosh computer under IgoPro (WaveMetrics). Data were analysed with Axograph (Axon Instruments). Spontaneous synaptic events were detected with a sliding-template algorithm (Axograph) with the threshold set at 3 – 4 s.d. Access resistance was 5 – $30 \text{ M}\Omega$ and was monitored online throughout the experiment. Experiments were rejected if series resistance changed by more than 10% . No series resistance compensation was used. Monosynaptic IPSCs were isolated by perfusion of $10 \mu\text{M}$ CNQX or 1 mM kynurenic acid. A monosynaptic IPSC, smaller in amplitude than the disynaptic IPSC, could be evoked in the presence of CNQX. In some instances the stimulus strength was then increased to improve the signal-to-noise ratio. Ten traces have been averaged for the figures.

All values are expressed as means $\pm$ s.e.m.. All statistical comparisons were done with Student's *t*-test. NHPP-spermine (N-(4-hydroxyphenyl)propanoyl)-spermine, Tocris Cookson) was made up freshly on the day of each experiment. Drugs used were CNQX, bicuculline methiodide (Tocris Cookson), D-APV (RBI Research Chemicals or Tocris Cookson) and picrotoxin (Sigma).

Received 2 April; accepted 15 June 1998.

- LeDoux, J. E. Emotion: clues from the brain. *Annu. Rev. Psychol.* **46**, 209–235 (1995).
- Davis, M., Rainnie, D. & Cassell, M. Neurotransmission in the rat amygdala related to fear and anxiety. *Trends Neurosci.* **17**, 208–214 (1994).

3. Rogan, M. T., Staubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**, 604–607 (1997).
4. Washburn, M. S., Numberger, M., Zhang, S. & Dingledine, R. Differential dependence on GluR2 expression of three characteristic features of AMPA receptors. *J. Neurosci.* **17**, 9393–9406 (1997).
5. Geiger, J. R. P. *et al.* Relative abundance of subunit mRNAs determines gating and Ca^{2+} permeability of AMPA receptors in principal neurons and interneurons in rat CNS. *Neuron* **15**, 193–204 (1995).
6. Quirk, G. J., Repa, C. & LeDoux, J. E. Fear conditioning enhances short-latency auditory responses of lateral amygdaloid neurons: parallel recordings in the freely behaving rat. *Neuron* **15**, 1029–1039 (1995).
7. Washburn, M. S. & Moises, H. C. Electrophysiological and morphological properties of rat basolateral amygdaloid neurons *in vitro*. *J. Neurosci.* **12**, 4066–4079 (1992).
8. Rainnie, D. G., Asprodingi, E. K. & Shinnick-Gallagher, P. Intracellular recordings from morphologically identified neurons of the basolateral amygdala. *J. Neurosci.* **69**, 1350–1362 (1993).
9. Connors, B. W. & Gutnick, M. J. Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.* **13**, 99–103 (1990).
10. McCormick, D. A., Connors, B. W., Lighthall, J. W. & Prince, D. A. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J. Neurophysiol.* **54**, 782–806 (1985).
11. Hestrin, S. Different glutamate receptor channels mediate fast excitatory synaptic currents in inhibitory and excitatory cortical neurons. *Neuron* **11**, 1083–1091 (1993).
12. Smith, B. N. & Dudek, F. E. Amino acid-mediated regulation of spontaneous synaptic activity patterns in the rat basolateral amygdala. *J. Neurophysiol.* **76**, 1958–1967 (1996).
13. Mahanty, N. K. & Sah, P. The physiology of excitatory synapses in the lateral and basolateral amygdala. *Soc. Neurosci. Abstr.* **22** (1996).
14. Sah, P., Hestrin, S. & Nicoll, R. A. Properties of excitatory postsynaptic currents in interneurons in rat hippocampus *in vitro*. *J. Physiol. (Lond.)* **430**, 605–616 (1990).
15. Geiger, J. R. P., Lubke, J., Roth, A., Frotscher, M. & Jonas, P. Submillisecond AMPA receptor-mediated signalling at a principal neuron–interneuron synapse. *Neuron* **18**, 1009–1023 (1997).
16. Farb, C. R., Aoki, C. & LeDoux, J. E. Differential localization of NMDA and AMPA receptor subunits in the lateral and basal nuclei of the amygdala: a light and electron microscopic study. *J. Comp. Neurol.* **362**, 86–108 (1995).
17. Hollmann, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
18. Jonas, P., Racca, C., Sakmann, B., Seeburg, P. H. & Monyer, H. Differences in Ca^{2+} permeability of AMPA-type glutamate receptor channels in neocortical neurons caused by differential GluR-B subunit expression. *Neuron* **12**, 1281–1289 (1994).
19. Jonas, P. & Burnashev, N. Molecular mechanisms controlling calcium entry through AMPA-type glutamate receptor channels. *Neuron* **15**, 987–990 (1995).
20. Racca, C., Catania, M. V., Monyer, H. & Sakmann, B. Expression of AMPA–glutamate receptor B subunit in rat hippocampal neurons. *Eur. J. Neurosci.* **8**, 1580–1590 (1996).
21. McBain, C. J. & Dingledine, R. Heterogeneity of synaptic glutamate receptors on CA3 stratum radiatum interneurons of rat hippocampus. *J. Physiol. (Lond.)* **462**, 373–392 (1993).
22. Isa, T., Itazawa, S., Iino, M., Tsuzuki, K. & Ozawa, S. Distribution of neurons expressing inwardly rectifying and Ca^{2+} -permeable AMPA receptors in rat hippocampal slices. *J. Physiol. (Lond.)* **491**, 719–733 (1996).
23. McDonald, A. J. Localization of AMPA glutamate receptor subunits in subpopulations of non-pyramidal neurons in the rat basolateral amygdala. *Neurosci. Lett.* **208**, 175–178 (1996).
24. Malenka, R. C., Kauer, J. A., Perkel, D. J. & Nicoll, R. A. The impact of postsynaptic calcium on synaptic transmission—its role in long term potentiation. *Trends Neurosci.* **12**, 444–450 (1989).
25. Rainnie, D. G., Asprodingi, E. K. & Shinnick-Gallagher, P. Inhibitory transmission in the basolateral amygdala. *J. Neurophysiol.* **66**, 999–1009 (1991).
26. Jia, Z. *et al.* Enhanced LTP in mice deficient in the AMPA receptor GluR2. *Neuron* **17**, 945–956 (1996).
27. Gu, J. G., Albuquerque, C., Lee, C. J. & MacDermott, A. B. Synaptic strengthening through activation of Ca^{2+} -permeable AMPA receptors. *Nature* **381**, 793–796 (1996).
28. Rogan, M. T. & LeDoux, J. E. LTP is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. *Neuron* **15**, 127–136 (1995).
29. Cobb, S. R., Buhl, E., Halasy, K., Paulsen, O. & Somogyi, P. Synchronization of neuronal activity in the hippocampus by individual GABAergic interneurons. *Nature* **378**, 75–78 (1995).
30. Blanton, M. G., Lo Turco, J. J. & Kriegstein, A. R. Whole cell recording from neurons in slices of reptilian and mammalian cerebral cortex. *J. Neurosci. Methods* **30**, 203–210 (1989).

Acknowledgements. We thank John Bekkers, Bob Callister and Troy Margrie for comments on the manuscript, and John Clements for help with analysis software in Axograph. P. Sah is a Charles and Sylvia Viertel Senior Medical Research Fellow. This work was supported by grants from the National Health and Medical Research Council of Australia.

Correspondence and requests for materials should be addressed to P.S. (e-mail: Pankaj.Sah@anu.edu.au).

Pore stoichiometry of a voltage-gated chloride channel

Christoph Fahlke, Thomas H. Rhodes, Reshma R. Desai & Alfred L. George Jr

Departments of Medicine (Nephrology) and Pharmacology, and The Center for Molecular Neuroscience, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-6600, USA

Ion channels allow ions to pass through cell membranes by forming aqueous permeation pathways (pores). In contrast to most known ion channels, which have single pores, a chloride channel belonging to the ClC family¹ (*Torpedo* ClC-0) has functional features that suggest that it has a unique ‘double-barrelled’ architecture in which each of two subunits forms an independent pore. This model is based on single-channel recordings of ClC-0

that has two equally spaced and independently gated conductance states^{2–4}. Other ClC isoforms do not behave in this way^{5,6}, raising doubts about the applicability of the model to all ClC channels. Here we determine the pore stoichiometry of another ClC isoform, human ClC-1, by chemically modifying cysteines that have been substituted for other amino acids located within the ClC ion-selectivity filter⁷. The ClC-1 channel can be rendered completely susceptible to block by methanethiosulphonate reagents when only one of the two subunits contains substituted cysteines. Thiol side chains placed at corresponding positions in both subunits can form intersubunit disulphide bridges and coordinate Cd^{2+} , indicating that the pore-forming regions from each subunit line the same conduction pathway. We conclude that human ClC-1 has a single functional pore.

We previously identified three amino-acid positions (Lys 231, Pro 234 and His 237) within the ion-selectivity filter (P1 region) of human ClC-1 (hClC-1) at which substituted thiol groups (designated K231C, P234C and H237C) projecting into the hClC-1 pore can be covalently modified by 2-(trimethylammonium)ethyl methanethiosulphonate (MTSET) or 2-sulphonatoethyl methanethiosulphonate (MTSES), resulting in channel block⁷. ClC channels are symmetrical dimers^{3,4,8,9}, and therefore the configuration of accessible pore-lining residues within the P1 region allows us to distinguish between quaternary architectures with one or two pores (Fig. 1). In a symmetrical double-barrelled channel, as has been suggested for ClC-0 (refs 2–4, 8), P1 regions from each subunit will project into two separate conduction pathways. In contrast, accessible P1 residues from each subunit will face the same aqueous cavity in a single-pore channel.

We expressed functional heterodimeric hClC-1 channels using tandem constructs⁹ in which only one of the two concatenated subunits has a P1-region cysteine substitution. If hClC-1 is a double-barrelled channel then modification of the heterodimer (Fig. 1a) by methanethiosulphonate should occur at the same rate but to a lesser extent than for a homodimeric cysteine-substituted channel (Fig. 1b). In contrast, if hClC-1 has only one pore lined by the P1 regions of both subunits, we would expect single cysteine-substituted heterodimeric channels (Fig. 1c) to be blocked to the same extent as the cysteine-substituted homodimer (Fig. 1d). Furthermore, we would expect that the time course of modification would be affected because of the reduced number of reactive thiols per pore and the change in the surrounding chemical environment¹⁰ (Fig. 1c).

Figure 2a illustrates the time course of whole-cell chloride currents recorded from cells expressing a heterodimeric K231C–K231A heterodimeric construct during modification by MTSES. We used the K231A mutant as the non-reactive subunit partner because it exhibits the same gating behaviour as K231C but is not susceptible to modification by MTS⁷. Extracellular application of

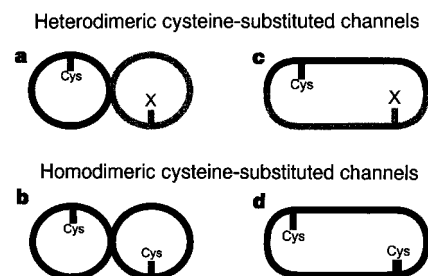


Figure 1 Possible configurations of hClC-1 channels with cysteine substitutions in pore-forming segments of one or both subunits. **a**, Double-barrelled channel in which only one subunit is cysteine substituted. **b**, Double-barrelled channel with cysteine substitution in both subunits. **c**, Single-barrelled channel with only one substituted cysteine. **d**, Single-barrelled channel with two substituted cysteines. X, any amino acid.

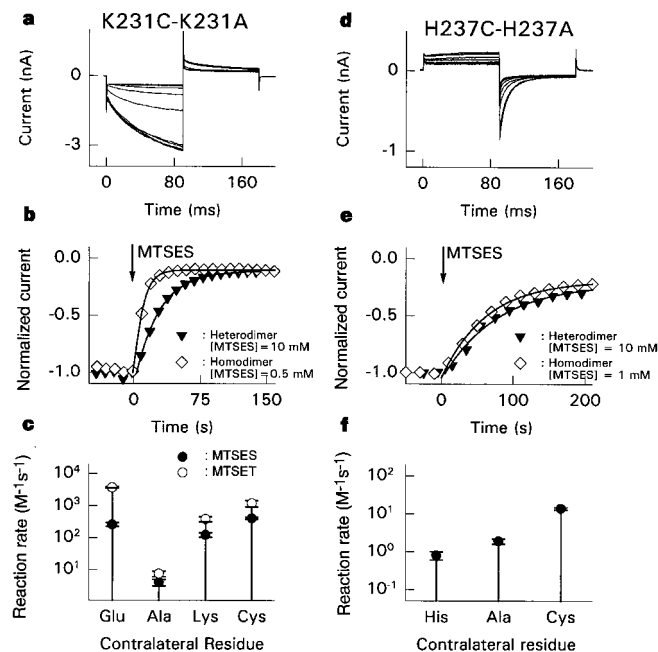


Figure 2 MTS-mediated modification of heterodimeric single-cysteine-substituted hClC-1 channels. **a**, Whole-cell current recordings during modification of K231C-K231A hClC-1 with 10 mM MTSES. The pulse protocol consisted of a voltage step to -105 mV followed by a $+75$ mV test potential. **b**, Time dependence of the normalized maximum current at -105 mV, obtained from the recordings shown in **a** and from a similar experiment performed with K231C homodimers using 0.5 mM MTSES. Because K231C homodimers are more reactive, we used a 20-fold lower MTSES concentration to resolve the reaction time course on the same plot as the less reactive heterodimeric channel. Solid lines represent mono-exponential fits. **c**, Second-order reaction rates obtained for the K231C homodimer (Cys) and three heterodimeric constructs: K231C-K231E (Glu), K231C-K231A (Ala) and K231C-wild-type (Lys). **d**, Excised inside-out patch-clamp recording during modification of H237C-H237A with 10 mM MTSES. The pulse protocol consisted of a voltage step to $+75$ mV followed by a -105 mV test potential. **e**, Time dependence of the maximum current at -105 mV, obtained from the recordings shown in **d** and from a similar experiment performed with H237C homodimers using a MTSES concentration of 1 mM. Solid lines represent mono-exponential fits. **f**, Second-order reaction rates obtained for three constructs: H237C-wild-type (His), H237C-H237A (Ala) and H237C (Cys).

10 mM MTSES to cells expressing K231C-K231A produced a rapid mono-exponential decay in current amplitude (Fig. 2b), but this reaction was much slower than that of homodimeric K231C treated with a 20-fold lower MTSES concentration (0.5 mM, Fig. 2b). The fraction of current remaining after achieving steady-state modification of K231C-K231A ($I_{\text{modification}}/I_{\text{control}} = 0.14 \pm 0.02$, $n = 7$) is not significantly different from the values observed for the K231C homodimer ($I_{\text{modified}}/I_{\text{control}} = 0.08 \pm 0.04$, $n = 4$) (Fig. 2b). These results indicate that the accessible amino-acid side chains from both subunits may be facing the same aqueous cavity.

The reaction rate in K231C-K231A is decreased much more than would be expected if the number of reactive thiols were reduced by half. Reaction rates are affected by the reactivity of the thiol group, which is dependent upon the thiol $pK^{10,11}$ and by the access pathway for MTS reagents¹². Both features can be altered by introducing a nearby charged residue. We therefore examined the time course of modification of K231C-X heterodimers by MTSES and MTSET (where X represents K231A, K231E or wild-type ClC-1). Figure 2c shows second-order rate constants for MTSET and MTSES modification of three different heterodimeric constructs in which the contralateral residue was anionic (K231C-K231E), neutral (K231C-K231A) or cationic (K231C-wild-type). These data indicate that the reaction rate constants are dependent on the identity of the P1 residue in the non-cysteine-substituted subunit. This result would be expected if the two corresponding residues face the same aqueous cavity.

Another P1 residue (His 237) located more deeply in the hClC-1 ion pore and accessible only to intracellular reagents can be tested similarly. For the heterodimeric H237C-H237A channel, application of MTSES to the cytoplasmic side of the membrane caused a mono-exponential decay in current (Fig. 2d, e) that was qualitatively similar to that observed after modification of H237C homodimers. The relative reduction of current amplitude for H237C-H237A ($I_{\text{modified}}/I_{\text{control}} = 0.20 \pm 0.01$, $n = 5$) is not significantly different from that observed after modification of H237C homodimers ($I_{\text{modified}}/I_{\text{control}} = 0.24 \pm 0.02$, $n = 4$). H237C-H237A exhibited a larger modification time constant than H237C homodimers exposed to a tenfold lower MTSES concentration, showing a reduction of MTSES reactivity in the heterodimer (Fig. 2e). Again, second-order reaction rate constants vary depending on the residue in the non-cysteine-substituted subunit (Fig. 2f). These data are consistent with the presence of a shared aqueous cavity at the

cytoplasmic face of the hClC-1 pore.

Two other experimental approaches show that P1 residues in separate subunits face the same pore. When cells expressing K231C homodimers were exposed to an oxidizing reagent, $\text{Cu(II)(1,10-phenanthroline)}_3$ (refs 13, 14), the current amplitude rapidly decreased with a mono-exponential time course to a relative amplitude of 0.07 ± 0.01 ($n = 6$). This process is fully reversible by application of dithiothreitol (DTT), which causes a mono-exponential increase in current amplitude back to near initial levels (Fig. 3a, b). Concentrations of $\text{Cu(II)(1,10-phenanthroline)}_3$ as low as 1.5 nM are sufficient to cause the full effect. Neither the extent of the block (Fig. 3c) nor the inverse time constants of current decay (Fig. 3d) depend on the reagent concentration when it is above 1.5 nM. Homodimeric K231A and heterodimeric K231C-K231A channels do not react in this way. High concentrations of $\text{Cu(II)(1,10-phenanthroline)}_3$ (150 μM) reduce current amplitudes of K231A homodimers and K231C-K231A heterodimers, but this effect is quite different from that observed with K231C homodimers. The current amplitude of the heterodimers is reduced by less than 50% (Fig. 3c); this reduction is not reversible with DTT treatment; and the remaining current shows an altered activation time course upon hyperpolarizing voltage steps. We conclude that the complete block observed after treatment of K231C homodimers with $\text{Cu(II)(1,10-phenanthroline)}_3$ depends on cysteine substitution at position 231 of both subunits and is due to intersubunit disulphide bridge formation. Because the thiol side chain of Cys 231 projects into the aqueous pore⁷, this disulphide bridge must connect both subunits by spanning the conduction pathway.

Single cysteine residues can bind Cd^{2+} individually or by coordinated ligation of the metal in conjunction with other nearby cysteines^{15,16}. Extracellular, but not intracellular, Cd^{2+} blocks wild-type hClC-1 (ref. 17), and so we tested only the accessibility of internal Cd^{2+} to cysteine residues introduced at positions 234 and 237 (ref. 7). Dose-response curves for intracellular Cd^{2+} -mediated block of wild-type, H237C-wild-type and H237C ClC-1 are shown in Fig. 4a, and the responses of P234A, P234C-P234A and P234C are shown in Fig. 4b. The dissociation constant (K_d) for Cd^{2+} block of H237C-wild-type (0.58 ± 0.1 mM) is nearly identical to the geometric mean (Fig. 4a, broken line) of the values for wild-type ClC-1 (15.5 ± 2.7 mM) and H237C (0.025 ± 0.004 mM), indicating that the free energy for Cd^{2+} binding varies linearly with the number of Cys 237 residues per channel. This suggests that Cd^{2+}

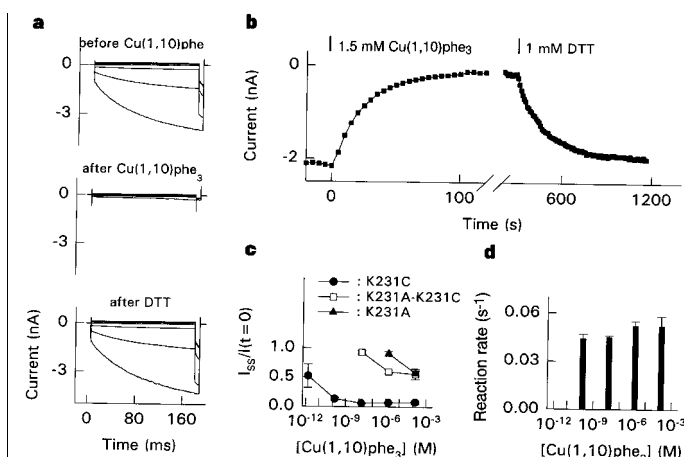


Figure 3 Formation of disulphide bridges between two Cys231 sidechains. **a**, Whole-cell recordings from a K231C-expressing cell before and after modification with $\text{Cu(II)(1,10-phenanthroline)}_3$, and after reduction by DTT. Traces were recorded during test potentials to voltages between -125 mV and $+75$ mV in 40 mV increments, each followed by a fixed -125 mV test pulse. **b**, Time course of current amplitude decrease upon application of $1.5 \mu\text{M}$ $\text{Cu(II)(1,10-phenanthroline)}_3$, and the subsequent current increase after application of 1 mM DTT. The maximum current recorded at -105 mV is plotted against time. **c**, Relative late current amplitude at -105 mV, after modification of three different constructs, plotted against $\text{Cu(II)(1,10-phenanthroline)}_3$ concentration. **d**, Inverse time constant of current reduction for different $\text{Cu(II)(1,10-phenanthroline)}_3$ concentrations.

interacts simultaneously with both Cys237 residues¹⁸. The K_d of the P234C–P234A heterodimer ($35.6 \pm 10.1 \mu\text{M}$) is greater than the geometric mean K_d (Fig. 4b, broken line) for unsubstituted P234A homodimers ($132.7 \pm 42.2 \mu\text{M}$) and cysteine-substituted P234C homodimers ($0.0016 \pm 0.0012 \mu\text{M}$). The two Cys234 residues cooperatively bind one Cd^{2+} , indicating that they are energetically coupled in the context of Cd^{2+} binding. This energetic coupling indicates that the two Cys234 residues may be in close proximity^{19,20}, in agreement with the earlier finding that the main constriction of the hCIC-1 pore is near to this residue⁷.

The observed high-affinity coordination of Cd^{2+} by two P1 cysteines is possible only if both residues face the same pore. If hCIC-1 were double-barrelled and each introduced cysteine provided an independent Cd^{2+} -binding site that resulted in the block of only one protopore, then the Cd^{2+} dose-response for single cysteine-substituted heterodimers would be a composite curve, equal to the sum of the dose-response curves of homodimeric cysteine-substituted and non-cysteine-substituted channels, weighted by their relative single-channel amplitudes. Alternatively, if both protopores could be blocked by a single Cd^{2+} through an allosteric interaction, then the dose-response curve of the heterodimeric channel would be a single Langmuir isotherm with a dissociation constant approximately twice that observed for a cysteine-substituted homodimer (Fig. 4, dotted lines, as calculated for two independent and distinct binding sites)¹⁸. Neither prediction is fulfilled by our data.

Using three independent experimental approaches, we have shown that the accessible residues within the P1 region of both hCIC-1 subunits face the same aqueous cavity. The P1 region spans the narrowest part of the CIC conduction pathway and contains elements involved in ion binding and selection⁷. Although it is possible that Lys231 and His237 reside in shared outer and inner vestibules connected by two separate protopores, these protopores would lack structural features necessary for ion binding. Moreover, the finding that the Cys234 residues of each subunit are in close proximity restricts the maximum possible length of these protopores to only two residues. We conclude that hCIC-1 has a single functional pore. What accounts for the apparent differences in the

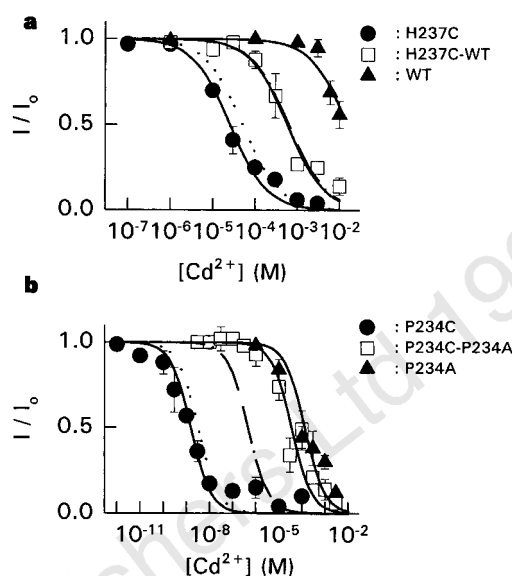


Figure 4 Dose-response curves for Cd^{2+} block. The relative current amplitudes recorded at $+75$ mV in different Cd^{2+} concentrations are shown for: **a**, wild-type, H237C-wild-type and H237C; and **b**, for P234C, P234C–P234A and P234A. Solid lines represent fits with Langmuir isotherms. Dotted lines represent calculated values for a heterodimeric channel with two independent and distinct binding sites for Cd^{2+} (calculated dissociation constants: H237C-wild-type, $50.9 \mu\text{M}$; P234C–P234A, 1.5 nM)¹⁸. The broken lines represent predicted values for a heterodimeric channel in which two different sites simultaneously interact with one Cd^{2+} ion¹⁸. Dose-response curves of a heterodimeric channels were simulated under the assumption either that there are two independent and distinct binding sites for Cd^{2+} or that two different sites interact simultaneously with one Cd^{2+} ion¹⁸. The curves are predicted to be single Langmuir isotherms. Using experimentally obtained K_d values for homodimeric cysteine-substituted (K_{2C}) or non-substituted (K_{0C}) channels, the K_d of the heterodimeric channel can be calculated for both conditions as follows: if there are two independent Cd^{2+} -binding sites, $K_d = 2K_{2C}K_{0C}/(K_{2C} + K_{0C})$; if two sites interact simultaneously with one Cd^{2+} ion, $K_d = \sqrt{K_{2C}K_{0C}}$.

functional pore stoichiometry of CIC-0 and hCIC-1? Perhaps CIC-0 and hCIC-1 truly differ in pore stoichiometry, despite their high degree of primary structure identity. Another possibility is that both CIC isoforms have two distinct pores formed by two structurally dissimilar hemipores contributed by each subunit, and that only one pore is functional in hCIC-1. However, this possibility violates the observed functional symmetry of CIC-0^{2–4,8}. Alternatively, CIC-0 could have only a single pore, and the characteristic single-channel behaviour of CIC-0, which originally led to formulation of the two-pore hypothesis, may have another explanation. Indeed, equally spaced and binomially distributed subconductance states have been described²¹ in a member of an ion-channel family that is firmly established to be single-barrelled^{22,23}. This shows that conformational changes in single pores may mimic the behaviour one expects for multiple ion-conduction pathways.

Methods

Mutagenesis and construction of heterodimers. We performed site-specific mutagenesis⁷, heterodimeric tandem construction⁹, and transient transfection of tsA201 cells²⁴ as described. Homodimeric channels were expressed by transfecting cells with monomeric constructs.

Electrophysiology. We performed standard whole-cell or inside-out patch-clamp recordings using experimental conditions as described⁷. The standard extracellular solution contained (in mM): NaCl 140, KCl 4, CaCl_2 2, MgCl_2 1, HEPES 5; the standard intracellular solution consisted of (in mM): NaCl 130, MgCl_2 2, EGTA 5, HEPES 10. All solutions were adjusted to pH 7.4.

Modification with MTS reagents was performed as described⁷. To promote formation of disulphide bridges, cells voltage-clamped in the whole-cell mode

were moved close to a perfusion pipette containing standard extracellular solution supplemented with $\text{Cu}(\text{II})(1,10\text{-phenanthroline})_3$ at different concentrations. A $\text{Cu}(\text{II})(1,10\text{-phenanthroline})_3$ stock solution was made by dissolving 150 mM CuSO_4 and 500 mM 1,10 phenanthroline in 4:1 water/ethanol. After reaching steady-state conditions, cells were moved to another perfusion pipette with standard extracellular solution containing 1 mM DTT. Current modifications under oxidizing and reducing conditions were observed as described for the MTS experiments. The time course was fit with a single exponential giving the time constant of modification.

To determine block by intracellular Cd^{2+} , Cd^{2+} was added in variable concentration to a modified intracellular solution devoid of EGTA containing (in mM): NaCl 140, MgCl_2 2, HEPES 10, pH 7.4. Initially, the voltage dependence of the instantaneous current amplitude after a 300-ms prepulse to +75 mV was obtained on inside-out patches in standard intracellular solution. Then the Cd^{2+} -containing solution was applied by moving the patch into the stream of a perfusion pipette. The time course of the block was determined by repetitive pulsing. After reaching a steady-state value, the voltage dependence of the instantaneous current amplitudes was measured with the patch pipette in the solution stream and shortly after switching to standard intracellular solution. Relative block was obtained by dividing the current amplitude at the end of the prepulse before and after Cd^{2+} application.

Data were analysed with a combination of PClamp (Axon Instruments) and SigmaPlot (Jandel Scientific) programs. All data are shown as means $\pm$ s.e.m.

Received 20 February; accepted 5 June 1998.

- Jentsch, T. J. Molecular biology of voltage-gated chloride channels. *Curr. Top. Membr.* **42**, 35–57 (1994).
- Miller, C. Open-state substructure of single chloride channels from *Torpedo* electroplax. *Phil. Trans. R. Soc. Lond. B* **299**, 401–411 (1982).
- Middleton, R. E., Pheasant, D. J. & Miller, C. Homodimeric architecture of a ClC-type chloride ion channel. *Nature* **383**, 337–340 (1996).
- Ludewig, U., Pusch, M. & Jentsch, T. J. Two physically distinct pores in the dimeric ClC-0 chloride channel. *Nature* **383**, 340–343 (1996).
- Malinowska, D. H., Kupert, E. Y., Bahinski, A., Sherry, A. M. & Cuppoletti, J. Cloning, functional expression, and characterization of a PKA-activated gastric Cl^- channel. *Am. J. Physiol. Cell Physiol.* **268**, C191–C200 (1995).
- Duan, D., Winter, C., Cowley, S., Hume, J. R. & Horowitz, B. Molecular identification of a volume-regulated chloride channel. *Nature* **390**, 417–421 (1997).
- Fahlke, Ch., Yu, H. T., Beck, C. L., Rhodes, T. H. & George, A. L. Jr Pore-forming segments in voltage-gated chloride channels. *Nature* **390**, 529–532 (1997).
- Middleton, R. E., Pheasant, D. J. & Miller, C. Purification, reconstitution, and subunit composition of a voltage-gated chloride channel from *Torpedo* electroplax. *Biochemistry* **33**, 13189–13198 (1994).
- Fahlke, Ch., Knittle, T. J., Gurnett, C. A., Campbell, K. P. & George, A. L. Jr Subunit stoichiometry of human muscle chloride channels. *J. Gen. Physiol.* **109**, 93–104 (1997).
- Roberts, D. D., Lewis, S. D., Ballou, D. P., Olson, S. T. & Shafer, J. A. Reactivity of small thiolate anions and cysteine-25 in papain toward methyl methanethiosulfate. *Biochemistry* **25**, 5595–5601 (1986).
- Fersht, A. R. *Enzyme Structure and Mechanism* (W. H. Freeman, New York, 1985).
- Cheung, M. & Akabas, M. H. Locating the anion-selectivity filter of the cystic fibrosis transmembrane conductance regulator (CFTR) chloride channel. *J. Gen. Physiol.* **109**, 289–299 (1997).
- Careaga, C. L. & Fahlke, J. J. Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor. Detection by disulfide trapping. *J. Mol. Biol.* **226**, 1219–1235 (1992).
- Benitah, J. *et al.* Molecular motions within the pore of voltage-dependent sodium channels. *Biophys. J.* **73**, 603–613 (1997).
- Glusker, J. P. Structural aspects of metal liganding to functional groups in proteins. *Adv. Protein Chem.* **42**, 1–76 (1991).
- Benitah, J., Tomaselli, G. F. & Marban, E. Adjacent pore-lining residues within sodium channels identified by paired cysteine mutagenesis. *Proc. Natl Acad. Sci. USA* **93**, 7392–7396 (1996).
- Rychkov, G. Y. *et al.* pH-dependent interactions of Cd^{2+} and a carboxylate blocker with the rat ClC-1 chloride channel and its R304E mutant in the Sf-9 insect cell line. *J. Physiol. (Lond.)* **501**, 355–362 (1997).
- Heginbotham, L. & MacKinnon, R. The aromatic binding site for tetraethylammonium ion on potassium channels. *Neuron* **8**, 483–491 (1992).
- Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38**, 835–840 (1984).
- Hidalgo, P. & MacKinnon, R. Revealing the architecture of a K^+ channel pore through mutant cycles with a peptide inhibitor. *Science* **268**, 307–310 (1995).
- Root, M. J. & MacKinnon, R. Two identical noninteracting sites in an ion channel revealed by proton transfer. *Science* **265**, 1852–1856 (1994).
- Li, M., Unwin, N., Stauffer, K. A., Jan, Y. N. & Jan, L. Y. Images of purified *Shaker* potassium channels. *Curr. Biol.* **4**, 110–115 (1994).
- Doyle, D. A. *et al.* The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Fahlke, Ch., Beck, C. L. & George, A. L. Jr A mutation in autosomal dominant myotonia congenita affects pore properties of the muscle chloride channel. *Proc. Natl Acad. Sci. USA* **94**, 2729–2734 (1997).

Acknowledgements. We thank L. DeFelice, P. Hidalgo, M. Holmgren, R. Horn, J. P. Johnson, N. Mitrovic and C. I. Petersen for discussions and review of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (Ch.F.), the Muscular Dystrophy Association (Ch.F., A.L.G.) and the National Institutes of Health (A.L.G.).

Correspondence and request for materials should be addressed to Ch.F. (e-mail: christoph.fahlke@mcmail.vanderbilt.edu).

A dual thrombin receptor system for platelet activation

Mark L. Kahn^{*†}, Yao-Wu Zheng^{*}, Wei Huang^{*}, Violeta Bigornia^{*}, Dewan Zeng^{*}, Stephen Moff^{*}, Robert V. Farese Jr[‡], Carmen Tam^{*} & Shaun R. Coughlin^{*†§}

^{*} Cardiovascular Research Institute and Daiichi Research Center, [†] Department of Medicine, [‡] Gladstone Institute of Cardiovascular Disease, and [§] Department of Cellular and Molecular Pharmacology, University of California, Box 0130, 505 Parnassus Avenue, San Francisco, California 94143-0130, USA

Platelet-dependent arterial thrombosis triggers most heart attacks and strokes. Because the coagulation protease thrombin is the most potent activator of platelets¹, identification of the platelet receptors for thrombin is critical for understanding thrombosis and haemostasis. Protease-activated receptor-1 (PAR1) is important for activation of human platelets by thrombin^{2–6}, but plays no apparent role in mouse platelet activation^{7–9}. PAR3 is a thrombin receptor that is expressed in mouse megakaryocytes¹⁰. Here we report that thrombin responses in platelets from PAR3-deficient mice were markedly delayed and diminished but not absent. We have also identified PAR4, a new thrombin-activated receptor. PAR4 messenger RNA was detected in mouse megakaryocytes and a PAR4-activating peptide caused secretion and aggregation of PAR3-deficient mouse platelets. Thus PAR3 is necessary for normal thrombin responses in mouse platelets, but a second PAR4-mediated mechanism for thrombin signalling exists. Studies with PAR-activating peptides suggest that PAR4 also functions in human platelets, which implies that an analogous dual-receptor system also operates in humans. The identification of a two-receptor system for platelet activation by thrombin has important implications for the development of antithrombotic therapies.

PAR3-deficient mice (Fig. 1) developed normally and showed no spontaneous bleeding. Haematocrit, platelet counts and bleeding times¹¹ were indistinguishable from those of wild-type mice (data not shown). However, thrombin responses in PAR3-deficient platelets were markedly abnormal (Fig. 2). Wild-type platelets secreted their granule contents and aggregated reliably to 1 nM thrombin. In contrast, PAR3-deficient platelets were virtually unresponsive to 1 and 3 nM thrombin; 10 nM thrombin elicited delayed responses from PAR3-deficient platelets (Fig. 2) and the level of secretion eventually achieved by PAR3-deficient platelets was decreased compared with that in wild-type platelets. Even at 30 nM thrombin, the secretion response in PAR3-deficient platelets was delayed, but the level of secretion ultimately reached was comparable to that seen in wild-type platelets (Fig. 2). Secretion and aggregation in response to U46619, an agonist of the thromboxane receptor, were indistinguishable in wild-type and PAR3-deficient platelets even at submaximal agonist concentrations (data not shown). Thus PAR3 is necessary for normal responsiveness to thrombin in mouse platelets.

However diminished, thrombin responses did persist in PAR3-deficient platelets. What mediates these responses? The prototypical thrombin receptor PAR1 (ref. 2) plays no role in activation of mouse platelets by thrombin⁹, and PAR3-deficient platelets, like wild-type platelets^{7–9}, did not respond to PAR1-activating peptide (data not shown). Persistent thrombin signalling in PAR3-deficient platelets therefore suggested the presence of an as yet uncharacterized thrombin receptor—perhaps another PAR family member. A GenBank BLAST search for PAR-related sequences revealed an EST (for expressed sequence tag) encoding an 11-amino-acid sequence that was 73% identical to the cognate sequence in PAR2. Because this region is conserved among PARs, a full-length complementary DNA was obtained. This cDNA (GenBank accession

number AF080215) encoded a 397-amino-acid protein, now designated PAR4, that was most closely related to human PAR3, with 30% amino-acid sequence identity. Proteases activate PARs by cleaving their amino-terminal exodomains to unmask a new amino terminus that then serves as a tethered ligand, binding intramolecularly to the body of the receptor to effect transmembrane signalling^{2,12,13}. Examination of PAR4's amino-terminal exodomain revealed a putative thrombin cleavage site (Fig. 3) that was identical to the thrombin cleavage site in rat PAR1 (ref. 14). Expression of PAR4 in *Xenopus* oocytes did confer robust signalling to thrombin, but, at least in this system, PAR4 required higher thrombin concentrations than did the well-studied thrombin receptor PAR1 (Fig. 3). Thrombin cleaved a Flag epitope fused to PAR4's amino terminus¹⁵ from the surface of PAR4-expressing oocytes, and active site-inhibited PPACK thrombin¹⁶ did not cause cleavage or signalling (data not shown), which is consistent with a proteolytic activation mechanism. PAR4 signalling was relatively thrombin-specific; among the related arginine/lysine-

specific proteases tested, only thrombin and trypsin elicited significant responses (Fig. 3).

Synthetic peptides that mimic the tethered ligands of PAR1 and PAR2 function as agonists for their respective receptors and have been used as pharmacological tools to probe the function of these receptors in various cell types^{2,5,6}; the cognate peptide for PAR3 appears to be insufficiently active to function as a free ligand¹⁰. The location of the putative thrombin cleavage site in PAR4 predicted that the sequence GYPGKF would serve as PAR4's tethered ligand, implying that a synthetic peptide of the same sequence might function as a PAR4 agonist. As predicted, this peptide activated PAR4-expressing *Xenopus* oocytes but not uninjected or water-injected oocytes (Fig. 3 and data not shown). In the thrombin receptors PAR1 and PAR3, the phenylalanine in the second position of the tethered ligand (a tyrosine in the PAR4-activating peptide) is critical for the function of this domain^{5,6,10}. Accordingly, the synthetic peptide GAPGKF had no activity at PAR4 (Fig. 3).

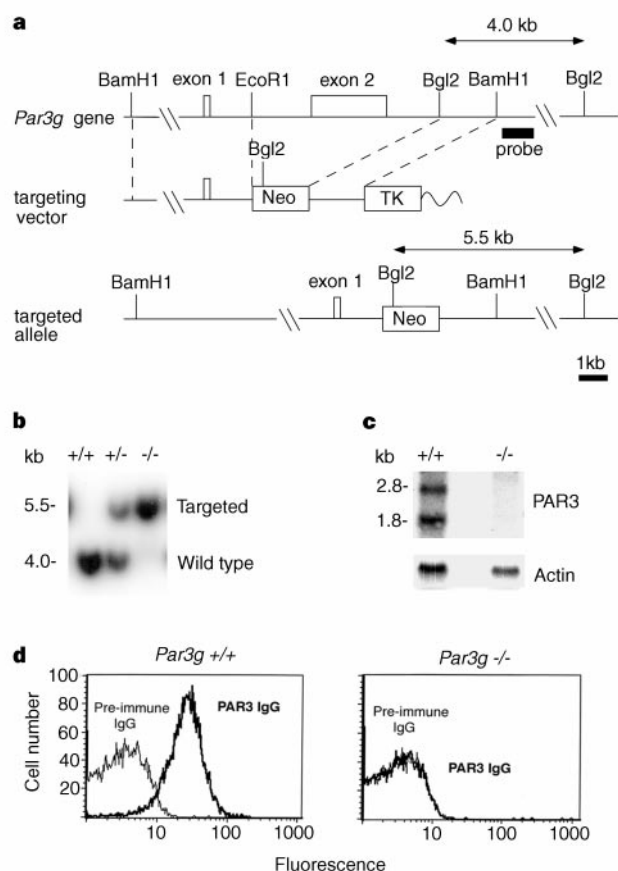


Figure 1 Generation of PAR3-deficient mice. **a**, Gene-targeting strategy. A replacement vector²⁵ was used to substitute a neomycin phosphotransferase expression cassette (Neo) for *Par3g* exon 2, which encodes the entire PAR3 protein except for its signal peptide. Wavy line represents plasmid backbone; TK, HSV thymidine kinase expression cassette. **b**, Southern-blot analysis of *Bgl*II-digested genomic DNA from the tails of pups derived from *Par3g*^{+/+} matings using 3' flanking probe (**a**). Targeting removed an endogenous *Bgl*II site and introduced a new *Bgl*II site. The 4.0-kb and 5.5-kb bands correspond to wild-type and targeted alleles, respectively. **c**, Northern-blot analysis of *Par3g*^{+/+} and *Par3g*^{-/-} mouse spleen mRNA using *Par3g* exon 2 probe and β -actin probe to control for lane loading. **d**, Flow cytometric analysis of wild-type and *Par3g*^{-/-} platelets for PAR3 protein. Platelets were incubated with preimmune IgG (narrow line) or PAR3 IgG (wide line). Bound IgG was detected with FITC-labelled 2° antibody²³. Each figure represents an analysis of 10,000 events. Note lack of surface PAR3 in *Par3g*^{-/-} platelets.

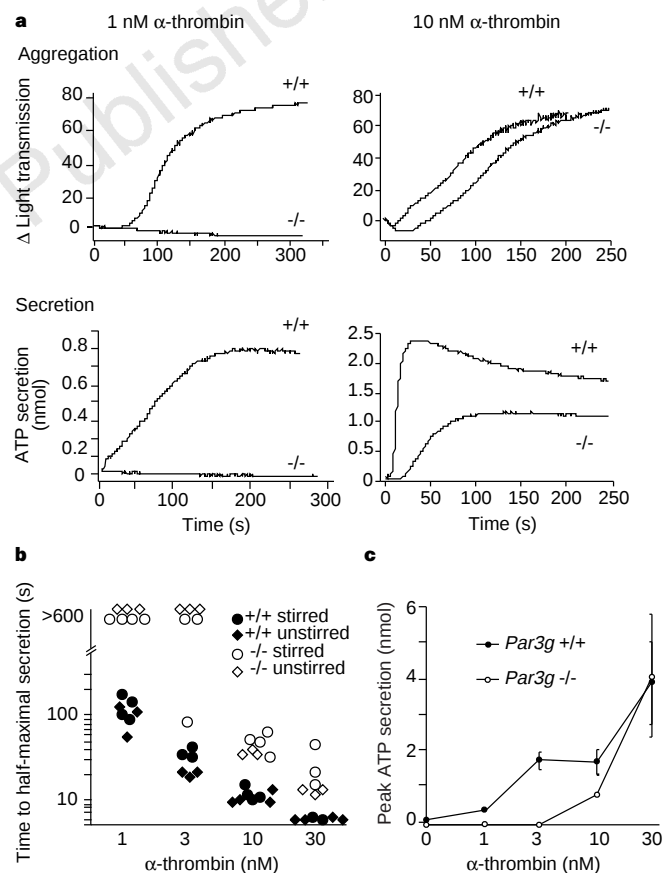


Figure 2 Thrombin responses in PAR3-deficient platelets. **a**, Aggregation and secretion of wild-type (*Par3g*^{+/+}) and PAR3-deficient (*Par3g*^{-/-}) platelets to 1 nM (left) and 10 nM (right) α -thrombin. Platelets were stirred and exposed to thrombin at 0 s. Similar results were obtained with five separate platelet preparations. **b**, Time to half-maximal secretion. *Par3g*^{+/+} (filled symbols) and *Par3g*^{-/-} (open symbols) platelets were stirred (circles) or left unstirred (diamonds) in a lumiaggregometer. Time to 50% of the peak response elicited by each of the indicated concentrations of thrombin was measured. Unstirred platelets did not aggregate. At each thrombin concentration, each point represents the response of a separate platelet preparation. Points at >600 s had no measurable secretion at 10 min. **c**, Dose response of *Par3g*^{+/+} (filled circles) and *Par3g*^{-/-} (open circles) platelets. Unstirred platelets were exposed to various concentrations of thrombin and peak ATP secretion was measured. Values are the mean responses ($\pm$ s.e.m.) of three separate platelet preparations. Some error bars are obscured by the symbols.

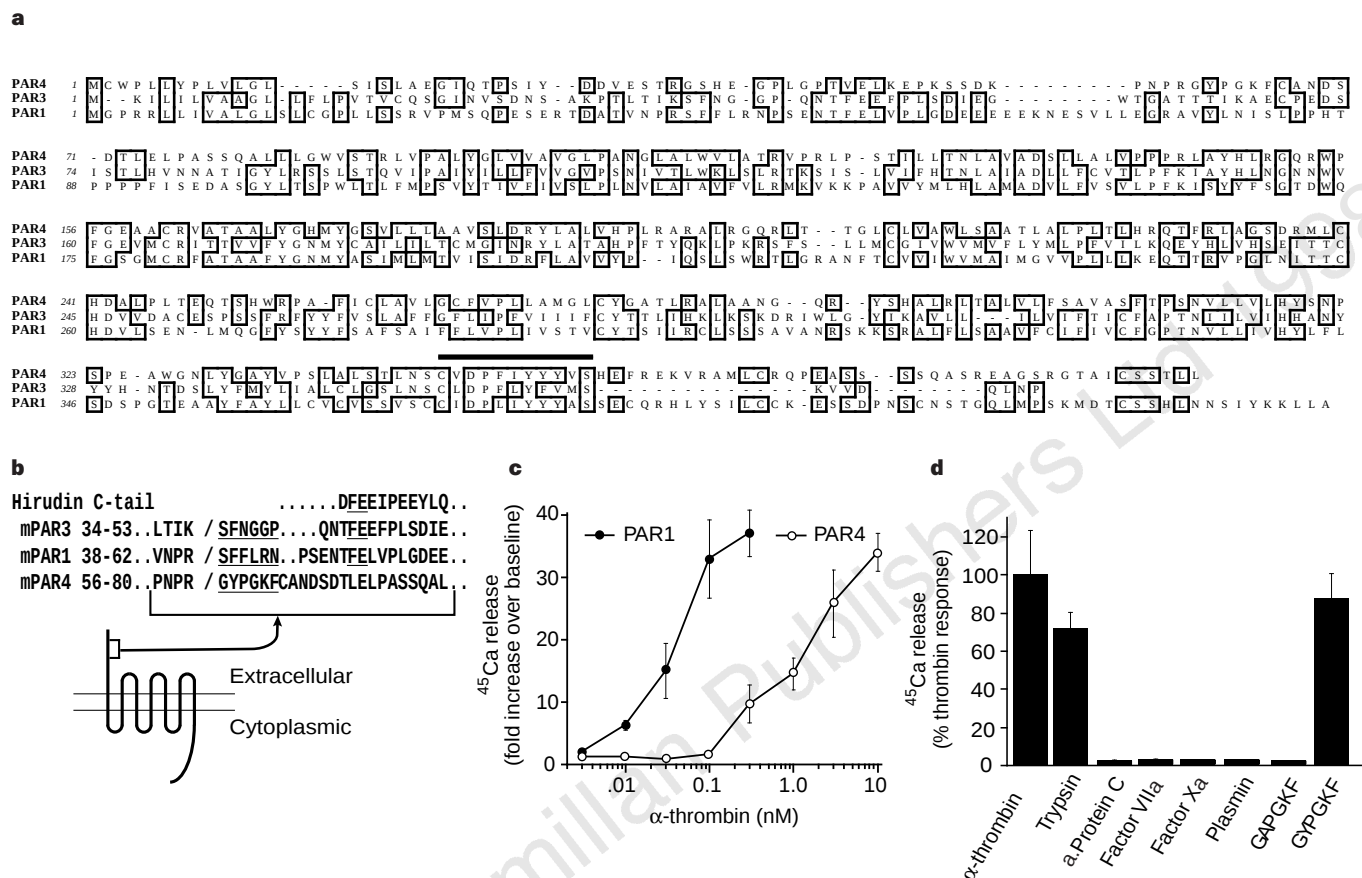


Figure 3 PAR4 amino-acid sequence and signalling properties. **a**, Sequence of PAR4 aligned with PAR3 and PAR1, the two known mouse thrombin receptors. Regions of amino-acid identity are boxed. The 11 amino acids of EST 400689 are overlined. **b**, Features of the amino-terminal exodomains of mouse thrombin receptors. The predicted thrombin cleavage site in PAR4 (I) is aligned with the known sites in PARs 1 and 3. The predicted or known tethered ligand sequences that are unmasked by receptor cleavage are underlined, as are amino-acid sequences resembling hirudin's carboxyl tail. The latter sequence in human PAR1 binds to thrombin's fibrinogen-binding exosite in a similar way to hirudin's C-tail^{12,26,27}. **c**, Thrombin activation of PAR4 and PAR1. Agonist-triggered ⁴⁵Ca release was measured in *Xenopus* oocytes expressing mouse PAR4 and human PAR1 tagged at their amino termini with a FLAG epitope¹⁵. Data are mean $\pm$ s.e.m.

Taken together, these data show that PAR4 is a functional protease-activated receptor that is capable of mediating thrombin signalling.

Northern-blot analysis of mouse tissues revealed a relatively high level of expression of PAR4 in spleen, trace expression in heart, lung, skeletal muscle and kidney, and no detectable expression in brain, liver or testis (data not shown). *In situ* hybridization of spleen and bone marrow revealed that, like PAR3 (ref. 10), PAR4 mRNA is expressed in mouse megakaryocytes (Fig. 4). Finally, GYPGKF, a PAR4-activating peptide, activated both wild-type and PAR3-deficient mouse platelets, whereas the mutant GAPGKF peptide had no activity (Fig. 4 and data not shown). These data suggest that PAR4 functions in both wild-type and PAR3-deficient mouse platelets and is likely to be responsible for the residual thrombin signalling seen in the latter.

The characterization of a dual thrombin receptor system in mouse platelets leads to the question of whether a similar system operates in human platelets. Unlike mouse platelets, thrombin signalling in human platelets is mediated at least in part by PAR1 (refs 2–6, 12). A role for PAR3 in human platelets has not been demonstrated, but pharmacological studies have suggested that PAR1-independent mechanisms for activating human platelets may exist^{3,17–20}. We have identified a human PAR4 cDNA (GenBank

(*n* = 4) and are expressed as the fold increase over baseline for each receptor. Surface expression of PAR4 measured with anti-FLAG monoclonal antibody was 1.8 times that of PAR1. Similar signalling data were obtained with untagged receptors. **d**, Specificity of PAR4 activation by proteases and peptides. ⁴⁵Ca release by oocytes expressing PAR4 was measured in response to the indicated active proteases (all at 5 nM except trypsin (0.5 nM)) or to the synthetic peptides GYPGKF and GAPGKF (each at 500 μ M). Uninjected and buffer-injected oocytes did not respond to proteases or peptides at these concentrations. Data are mean $\pm$ s.e.m. (*n* = 3) and are expressed as a percentage of the maximum response to thrombin. In the experiment shown, thrombin caused a 33-fold increase in ⁴⁵Ca release. These experiments were replicated at least twice.

accession number AF080214) and studies using polymerase chain reaction (PCR) after reverse transcription of RNA suggest that it is expressed in human platelets (M.L.K., M. Nakanishi-Matsui and S.R.C., manuscript in preparation). In addition, both SFLLRN, a PAR1-activating peptide, and GYPGKF, a PAR4-activating peptide, activated human platelets; the mutant peptide GAPGKF did not (Fig. 4). The potency of these pharmacological tools for platelet activation correlated with their potency at their respective receptors in the oocyte system. GYPGKF activated human platelets at a threshold concentration of 300–500 μ M and PAR4-expressing *Xenopus* oocytes with an EC₅₀ (effector concentration for a half-maximal response) of 30 μ M. SFLLRN activated human platelets at a threshold concentration of 1–3 μ M and PAR1-expressing oocytes with an EC₅₀ of 0.3 μ M. SFLLRN showed no activity on PAR4-expressing oocytes and GYPGKF showed little activity on PAR1-expressing oocytes (<8% of maximal PAR1 activation even at 500 μ M GYPGKF, data not shown). Lastly, GYPGKF activated PAR1-desensitized platelets (Fig. 4). These data suggest that PAR4 functions in human platelets.

In summary, PAR3 is necessary for normal thrombin signalling in mouse platelets but PAR4, a newly characterized thrombin receptor, also contributes. Human platelets also seem to use at least two

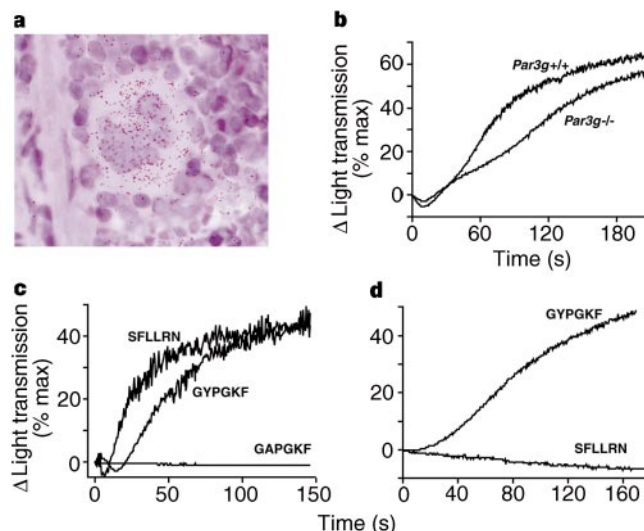


Figure 4 Expression of PAR4 in mouse megakaryocytes and evidence for PAR4 function in mouse and human platelets. **a**, *In situ* hybridization of mouse spleen for PAR4 mRNA. Bright-field photomicrograph shows silver grains overlying a megakaryocyte. Sense-probe controls were negative. **b**, GYPGKF, a PAR4-activating peptide, activates wild-type and PAR3-deficient mouse platelets. Stirred wild-type and PAR3-deficient platelets were exposed to GYPGKF (500 μ M) at 0 s and aggregation was measured. Under the same conditions, no response to the control peptide GAPGKF was obtained (data not shown). This experiment was repeated three times. **c**, PAR4-activating peptide activates

human platelets. Naive human platelets were exposed to the PAR1 agonist SFLLRN (3 μ M), the PAR4 agonist GYPGKF (500 μ M) or the mutant peptide GAPGKF (500 μ M) and aggregation was measured. **d**, PAR4-activating peptide activates PAR1-desensitized human platelets. SFLLRN-desensitized platelets (see Methods) were exposed to either SFLLRN (500 μ M) or GYPGKF (500 μ M) and aggregation was measured. This experiment was replicated three times using platelets from two different donors. Note that GYPGKF activates naive and PAR1-desensitized human platelets.

receptors, PAR1 and PAR4, for thrombin signalling. Why might a two-receptor system for platelet activation by thrombin exist? It may simply provide redundancy in a pathway important for haemostasis. More interestingly, it may provide a mechanism for responding to proteases other than thrombin or to thrombin itself over a wider range of concentrations, for signalling with different tempos, for activating distinct downstream effectors, or for allowing differential regulation of receptor levels or function. The identification of this receptor system provides a framework for further defining the roles and relative importance of distinct PARs in platelets and other cells and for refining strategies for pharmaceutical development. For example, it may be necessary to block both PAR1 and PAR4 in human platelets to achieve an antithrombotic effect. Alternatively, the existence of a second receptor may provide a useful margin of safety for such potentially powerful therapeutic agents and/or the ability to block selected *in vivo* responses to thrombin.

Note added in proof: While this manuscript was in press, the identification of human PAR4 was reported independently²⁸. □

Methods

Inactivation of the gene encoding PAR3 (*Par3g*). A bacterial artificial chromosome (BAC) that contained *Par3g* was obtained by PCR screen of a 129/SvJ mouse genomic library (Genome Systems). A 6.5-kilobase (kb) *Bam*H1/*Eco*R1 fragment 5' of exon 2 and a 2.0-kb *Bgl*II/*Bam*H1 fragment 3' of exon 2 were cloned into the pNTK vector²¹ to create the targeting vector (Fig. 1a). A 0.8-kb *Nde*I fragment of *Par3g* 3' of the short arm of homology was used as a probe to identify both the wild-type and targeted alleles (Fig. 1a). RF8 ES cells²² (129/SvJae) were electroporated with the targeting construct and clones that were resistant to G418 and FIAU were selected and screened by Southern blot. A highly chimaeric male mouse that was derived using *Par3g*^{+/-} ES cells was bred to C57BL/6 females to generate approximately 50 F₁ *Par3g*^{+/-} mice. All experiments reported here were performed using the F₂ offspring of these mice.

Northern blot analysis. Poly(A)⁺ RNA (3 μ g) from wild-type and *Par3g*^{-/-} mouse spleens was electrophoresed on a denaturing gel, transferred to reinforced nitrocellulose membrane (Schleicher & Schuell), and hybridized to a

600-base-pair (bp) *Eco*R1/*Sal*I PAR3 exon 2 probe under high stringency.

Flow cytometry. Washed mouse platelets were resuspended in platelet buffer (20 mM Tris-HCl pH 7.4, 140 mM NaCl, 2.5 mM KCl, 1 mM MgCl₂, 1 mg ml⁻¹ glucose, 0.5% BSA, 1 μ M PGE₁ and 5 mM EDTA), incubated with 10 μ g ml⁻¹ of anti-PAR3 IgG²³ at 4 °C for 1 h, washed, and then incubated with FITC-conjugated goat anti-rabbit IgG (Molecular Probes) at 4 μ g ml⁻¹ for 30 min. Platelets were then washed three times and analysed by flow cytometry.

Platelet aggregation and secretion. Blood was collected into citrate buffer from the inferior vena cava of pentobarbital-anaesthetized mice. Blood from 3–4 *Par3g*^{-/-} mice or their wild-type littermates was pooled for each platelet study. Platelet-rich plasma was prepared by centrifugation of whole blood at 200 g for 7 min. EDTA (10 mM) and PGE₁ (1 μ M) were then added and platelet-rich plasma was centrifuged at 500 g for 10 min. Platelets were then washed in platelet buffer containing 1 mM EDTA and 1 μ M prostaglandin PGE₁, collected by centrifugation, resuspended to an OD₅₀₀ of 1.0 (~2.5 × 10⁸ platelets per millilitre) in platelet buffer lacking EDTA and PGE₁, and incubated on ice for 30 min before use. Calcium chloride was added to a final concentration of 2 mM and aggregation and secretion were measured in a Chrono-Log lumiaggregometer. Platelet suspension (300 μ l) was added to the aggregometer chamber and change in light transmission after addition of agonist was followed. Results are expressed as the change (Δ) in light transmission, defined as the per cent increase in light transmission over that of the unactivated platelet suspension, with 100% representing light transmission of platelet buffer alone. Platelet ATP secretion was measured by adding luciferase (880 units per millilitre) and luciferin (8 μ g ml⁻¹) to each sample; the luminescence generated by platelet-released ATP was compared with that of an ATP standard. Human blood was drawn from the antecubital vein; otherwise human platelets were prepared in a manner identical to mouse platelets. For PAR1-desensitization studies, human platelets that were resuspended from the first platelet pellet were incubated with 100 μ M SFLLRN peptide at room temperature for 5 min without stirring, and then washed and resuspended as above.

Cloning of PAR4. Mouse EST 400689 was identified by BLAST search of the NCBI EST database using the coding region of human PAR2. The EST was sequenced and found to encode the carboxy-terminal 32 amino acids of a presumed G-protein-coupled receptor. A full-length cDNA was obtained by a combination of 5' rapid amplification of cloned ends (RACE) using cDNA

from mouse embryo at day 14–15 (Marathon cDNA, Clontech), and PCR of cDNA from a mouse brain endothelial cell line (bEND cells, W. Risau). A BAC mouse genomic clone was obtained from Genome Systems. The sequence of the cDNA (Fig. 3) and genomic clones were consistent. Human PAR4 was cloned from K562 mRNA by RT-PCR using primers based on mouse PAR4 sequence.

Functional studies in *Xenopus* oocytes. cDNA encoding Flag-epitope-tagged PAR4 (ref. 15) was constructed such that the Flag epitope was followed by amino-acid 18 of PAR4. Epitope-tagged and wild-type PAR4 cDNAs in pFROG² were transcribed *in vitro* and *Xenopus* oocytes were microinjected with 0.5–5.0 ng of PAR4 cRNA and 25 ng of PAR1 cRNA per oocyte. Agonist-triggered ⁴⁵Ca mobilization, a reflection of phosphoinositide hydrolysis in these cells, was measured^{2,24}.

In situ hybridization. *In situ* hybridization¹⁰ was done using sense or antisense ³⁵S-riboprobe transcribed from mouse PAR4 cDNA in pBluescript II SK. An 8-week exposure is shown.

Received 13 March; accepted 15 June 1998.

- Davey, M. & Luscher, E. Actions of thrombin and other coagulant and proteolytic enzymes on blood platelets. *Nature* **216**, 857–858 (1967).
- Vu, T.-K. H., Hung, D. T., Wheaton, V. I. & Coughlin, S. R. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* **64**, 1057–1068 (1991).
- Hung, D. T., Vu, T.-K. H., Wheaton, V. I., Ishii, K. & Coughlin, S. R. The cloned platelet thrombin receptor is necessary for thrombin-induced platelet activation. Blocking antiserum to the thrombin receptor's hirudin-like domain. *J. Clin. Invest.* **89**, 1350–1353 (1992).
- Brass, L. F. *et al.* Structure and function of the human platelet thrombin receptor. Studies using monoclonal antibodies directed against a defined domain within the receptor N terminus. *J. Biol. Chem.* **267**, 13795–13798 (1992).
- Vassallo, R. J., Kieber, E. T., Cichowski, K. & Brass, L. F. Structure-function relationships in the activation of platelet thrombin receptors by receptor-derived peptides. *J. Biol. Chem.* **267**, 6081–6085 (1992).
- Scarborough, R. M. *et al.* Tethered ligand agonist peptides. Structural requirements for thrombin receptor activation reveal mechanism of proteolytic unmasking of agonist function. *J. Biol. Chem.* **267**, 13146–13149 (1992).
- Connolly, T. M. *et al.* Species variability in platelet and other cellular responsiveness to thrombin receptor-derived peptides. *Thromb. Hemost.* **72**, 627–633 (1994).
- Derian, C. K., Santulli, R. J., Tomko, K. A., Haertlein, B. J. & Andrade-Gordon, P. Species differences in platelet responses to thrombin and SFLLRN. Receptor-mediated calcium mobilization and aggregation and regulation by protein kinases. *Thromb. Res.* **6**, 505–519 (1995).
- Connolly, A. J., Ishihara, H., Kahn, M. L., Farese, R. V. & Coughlin, S. R. Role of the thrombin receptor in development and evidence for a second receptor. *Nature* **381**, 516–519 (1996).
- Ishihara, H. *et al.* Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* **386**, 502–506 (1997).
- Novak, E. K. *et al.* Cocoa: a new mouse model for platelet storage pool deficiency. *Br. J. Haematol.* **69**, 371–378 (1988).
- Vu, T.-K. H., Wheaton, V. I., Hung, D. T. & Coughlin, S. R. Domains specifying thrombin-receptor interaction. *Nature* **353**, 674–677 (1991).
- Chen, J., Ishii, M., Wang, L., Ishii, K. & Coughlin, S. R. Thrombin receptor activation: confirmation of the intramolecular tethered ligand hypothesis and discovery of an alternative intermolecular ligand mode. *J. Biol. Chem.* **269**, 16041–16045 (1994).
- Zhong, C., Hayzer, D. J., Corson, M. A. & Runge, M. S. Molecular cloning of the rat vascular smooth muscle thrombin receptor. Evidence for *in vitro* regulation by basic fibroblast growth factor. *J. Biol. Chem.* **267**, 16975–16979 (1992).
- Ishii, K., Hein, L., Kobilka, B. & Coughlin, S. R. Kinetics of thrombin receptor cleavage on intact cells: relation to signalling. *J. Biol. Chem.* **268**, 9780–9786 (1993).
- Kettner, C. & Shaw, E. D-phe-pro-arg-CH₂Cl: a selective affinity label for thrombin. *Thromb. Res.* **14**, 969–973 (1979).
- Shuman, M., Botney, M. & Fenton, J. I. Thrombin-induced platelet secretion. *J. Clin. Invest.* **63**, 1211–1218 (1979).
- Goodwin, C. A. *et al.* Thrombin receptor activating peptide does not stimulate platelet procoagulant activity. *Biochem. Biophys. Res. Commun.* **202**, 321–327 (1994).
- Kramer, R. M. *et al.* Differential activation of cytosolic phospholipase A2 (cPLA2) by thrombin and thrombin receptor agonist peptide in human platelets. Evidence for activation of cPLA2 independent of the mitogen-activated protein kinases ERK1/2. *J. Biol. Chem.* **270**, 14816–14823 (1995).
- Henriksen, R. A., Samokhin, G. P. & Tracy, P. B. Thrombin-induced thromboxane synthesis by human platelets. Properties of anion binding exosite I-independent receptor. *Arterioscler. Thromb. Vasc. Biol.* **17**, 3519–3526 (1997).
- Mortensen, R. in *Current Protocols in Molecular Biology* (ed. Ausubel, F. M.) 9.16.1 (Wiley, New York, 1993).
- Meiner, V. L. *et al.* Disruption of the acyl-CoA:cholesterol acyltransferase gene in mice: evidence suggesting multiple cholesterol esterification enzymes in mammals. *Proc. Natl Acad. Sci. USA* **93**, 14041–14046 (1996).
- Ishihara, H., Zeng, D., Connolly, A. J., Tam, C. & Coughlin, S. R. Antibodies to protease-activated receptor 3 inhibit activation of mouse platelets by thrombin. *Blood* **91**, 4152–4157 (1998).
- Williams, J. A., McChesney, D. J., Calayag, M. C., Lingappa, V. R. & Logsdon, C. D. Expression of receptors for cholecystokinin and other Ca²⁺-mobilizing hormones in *Xenopus* oocytes. *Proc. Natl Acad. Sci. USA* **85**, 4939–4943 (1988).
- Capecchi, M. R. Altering the genome by homologous recombination. *Science* **244**, 1288–1292 (1989).
- Liu, L., Vu, T.-K. H., Esmon, C. T. & Coughlin, S. R. The region of the thrombin receptor resembling hirudin binds to thrombin and alters enzyme specificity. *J. Biol. Chem.* **266**, 16977–16980 (1991).
- Mathews, I. L. *et al.* Crystallographic structures of thrombin complexed with thrombin receptor peptides: existence of expected and novel binding modes. *Biochemistry* **33**, 3266–3279 (1994).
- Xu, W. F. *et al.* Cloning and characterization of human protease-activated receptor 4. *Proc. Natl Acad. Sci. USA* **95**, 6642–6646 (1998).

Acknowledgements. We thank T. Yu for blastocyst injection and H. Bourne for critical reading of this manuscript and S. E. Millar for assistance with figures.

Correspondence and requests for materials should be addressed to S.R.C. (e-mail: shaun_coughlin@quickmail.ucsf.edu).

A mutation in succinate dehydrogenase cytochrome *b* causes oxidative stress and ageing in nematodes

Naoaki Ishii*, Michihiko Fujii†, Philip S. Hartman‡, Michio Tsuda*, Kayo Yasuda*, Nanami Senoo-Matsuda*§, Sumino Yanase*, Dai Ayusawa† & Kenshi Suzuki*

* Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan

† Department of Biochemistry, Kihara Institute for Biological Research, Yokohama City University, Totsuka, Yokohama, Kanagawa 244-0813, Japan

‡ Department of Biology, Texas Christian University, Fort Worth, Texas 76129, USA

§ Tokyo Research Laboratory, Kyowa Hakko Kogyo Co. Ltd., Machida, Tokyo 194-8533, Japan

Much attention has focused on the aetiology of oxidative damage in cellular and organismal ageing^{1–4}. Especially toxic are the reactive oxygen byproducts of respiration and other biological processes⁵. A *mev-1(kn1)* mutant of *Caenorhabditis elegans* has been found to be hypersensitive to raised oxygen concentrations^{6,7}. Unlike the wild type, its lifespan decreases dramatically as oxygen concentrations are increased from 1 to 60% (ref. 7). Strains bearing this mutation accumulate markers of ageing (such as fluorescent materials and protein carbonyls) faster than the wild type^{8,9}. We show here that *mev-1* encodes a subunit of the enzyme succinate dehydrogenase cytochrome *b*, which is a component of complex II of the mitochondrial electron transport chain. We found that the ability of complex II to catalyse electron transport from succinate to ubiquinone is compromised in *mev-1* animals. This may cause an indirect increase in superoxide levels, which in turn leads to oxygen hypersensitivity and premature ageing. Our results indicate that *mev-1* governs the rate of ageing by modulating the cellular response to oxidative stress.

Three-factor crosses using visible genetic markers placed *mev-1* between *unc-50(e306)* and *unc-49(e382)* on chromosome III (Fig. 1). We tested cosmid from this region for their ability to rescue *mev-1* mutants from oxygen hypersensitivity after germline transformation¹⁰. Cosmid T07C4, but not C38H2, M03C11 or other cosmid, was able to rescue the *mev-1* mutant phenotype (Fig. 1). By testing various subclones from this cosmid, we identified a 5.6-kilobase (kb) fragment that also restored wild-type resistance (Figs 1, 2). Rescue was essentially complete with respect to both oxygen hypersensitivity (Fig. 2a) and premature ageing (Fig. 2b). This fragment includes an open reading frame containing the conceptual gene *cyt-1*, which is homologous to the bovine succinate dehydrogenase (SDH) cytochrome *b*₅₆₀ (ref. 11; GenBank accession number L26545 for *C. elegans* and S74803 for bovine, respectively). We found that the *mev-1(kn1)* strain contained a missense mutation resulting in a glycine-to-glutamic acid substitution in *cyt-1* (Fig. 3). This mutation created a restriction fragment-length polymorphism that enabled the restriction enzyme *MroI* to cleave wild-type DNA but not *mev-1* DNA at position 323. As predicted from the sequences, digestion by *MroI* of the products of polymerase chain reaction with reverse transcription (RT-PCR) yielded two bands from wild type (N2), one band from *mev-1(kn1)* and three bands from transgenic animals (*kn1;knIs2*) (Fig. 1). This confirms that the wild-type *cyt-1* gene that we introduced into the *mev-1* strain was expressed and that rescue was achieved by this gene. We also confirmed that a 2.8-kb wild-type PCR product of this gene, including regions from 1,566 bases upstream and 537 bases downstream, rescued *mev-1* (Fig. 1).

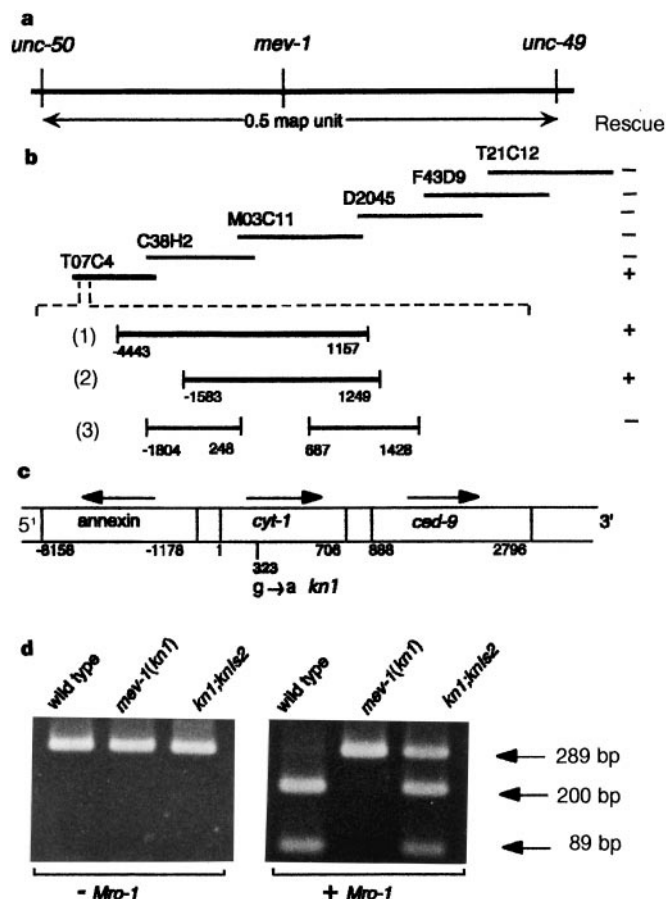


Figure 1 Genetic, physical and functional maps of the *mev-1* region. **a**, *mev-1* was positioned by a 3-factor cross: *unc-50* (2/4) *mev-1* (2/4) *unc-49*. **b**, Cosmid and DNA fragment rescue of the *mev-1* oxygen-sensitivity phenotype: rescue was determined by finding the surviving roller adults after exposing *mev-1* animals co-injected with *rol-6* gene and cosmids or DNA fragments to 90% oxygen in an airtight plastic chamber for four days. A plus sign indicates rescue and a minus sign indicates no rescue. (1) 5.6-kb fragment in cosmid T07C4; (2) 2.8-kb PCR product generated from wild-type animals; (3) a mixture of two wild-type PCR products. **c**, Gene structure around the rescued DNA fragments. The arrows show the direction of transcription. The numbers on the 5.6-kb fragment, wild-type PCR products and gene structure indicate the distances from the initiation codon in *cyt-1*. The location of the g to a transition in *kn1* is indicated. **d**, Agarose gel depicting undigested and *Mro1*-digested *cyt-1* RT-PCR products isolated from wild-type (N2) (lane 1), *mev-1*(*kn1*) (lane 2) and transgenic animals transformed with the 5.6-kb fragment (*kn1*;*knls2*) (lane 3).

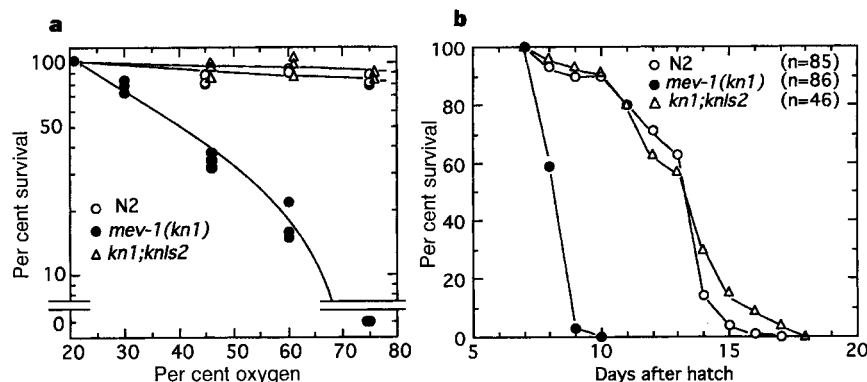


Figure 2 Comparison of oxygen sensitivity and lifespans of wild-type (N2), *mev-1* (*kn1*) and rescued transgenic animals (*kn1*;*knls2*). **a**, Survival under various oxygen concentrations. Survival at each point was relative to that of the same strain under atmospheric conditions. Between 50 to 800 animals were scored for

Table 1 Activity of SDH and complex II in wild-type and *mev-1* extracts

	Activities (nmol min ⁻¹ mg ⁻¹ protein)	
	Wild type	<i>mev-1</i>
Membrane-bound SDH	134 ± 21 (n = 3)	128 ± 11 (n = 3)
Complex II in membrane	54 ± 13 (n = 3)	5.1 ± 1.1 (n = 3)
Free SDH in cytosol	0 (n = 2)	0 (n = 2)

n, Number of experiments.

Although *cyt-1* and *ced-9*, a negative regulator of apoptosis in *C. elegans*, share a common promoter¹², results from the following experiments indicate that this juxtaposition carries no functional significance. First, a *mev-1*(*kn1*)/*ced-9*(*n1950n2161*) heterozygote was constructed. Although brood sizes were reduced in these *trans* double heterozygotes, they showed neither abnormal apoptosis nor oxygen hypersensitivity (data not shown). Thus, as also indicated by the transformation rescue data, these results indicate that *mev-1* and *ced-9* are not allelic. Second, neither *ced-9* gain-of-function nor loss-of-function homozygotes were hypersensitive to oxygen (data not shown).

Electron transport is mediated by five multimeric complexes (complexes I–V) that are embedded in the inner membrane of the mitochondrion. Mitochondrial succinate–ubiquinone reductase (complex II) catalyses electron transport from succinate to ubiquinone and is composed of succinate dehydrogenase (SDH), flavin protein, iron–sulphur protein and two other subunits containing cytochrome *b*₅₆₀. *In vivo*, SDH is anchored to the inner membrane with the cytochrome *b*₅₆₀ (ref. 13) and is the catalytic component of complex II. It is possible to quantify specifically both SDH activity and complex II activity by using separate assays, which we did after wild-type and *mev-1* extracts were differentially centrifuged to separate mitochondria and mitochondrial membranes from the cytosol (Table 1). We found that SDH activity in the *mev-1* mitochondrial fraction was identical to that of wild type, whereas complex II activity in the *mev-1* membrane fraction was reduced by more than 80% relative to wild type. As expected for a mitochondrial enzyme, there was no SDH activity in the cytosol. Thus, the *mev-1* mutation affects neither SDH anchoring to the membrane nor SDH activity *per se*, but it dramatically compromises the ability of complex II to participate in electron transport: this indicates that cytochrome *b*₅₆₀ (that is, the *mev-1*/*cyt-1* gene product) participates directly in transporting electrons from SDH to CoQ. As cytochrome *b*₅₆₀ has three predicted membrane-spanning domains¹¹, the substitution of glutamic acid for glycine at position 71 (only two amino acids removed from a histidine residue (His 73) that is thought to serve as a haem ligand^{11,14}) could affect the ability of iron to accept and relinquish electrons, thereby explaining the complex II deficiency in the *mev-1* mutant.

each point. **b**, Lifespans of animals reared under 60% of oxygen at 25°C. The mean lifespans with standard deviations were 13.0 ± 2.0 for N2, 8.5 ± 0.6 for *mev-1*(*kn1*) and 13.4 ± 2.3 days for *kn1*;*knls2*, respectively. n, Number of animals.

<i>C. elegans</i>	1	1	---INIPTAILRGRSISRSFGTSIVTKSEFTPIQKGEYLLQ	47
Bovine	1	1	MAALLLRHYGRG--LRAELSPOLCIRHAYPLCPAIFETIERF--SS---RN	45
Human	1	1	MAALLLRHYGRG--LRAELSPOLCIRHAYPLCPAIFETIERF--RN---RN	45
E (kn1)				
<i>C. elegans</i>	48	RSK	RRRIAPLVLVQPOPTWMLGFRNISCVML--KTLVVGIGFALFF	96
Bovine	46	TTL	NPFLPPLSTGNSLPIHRTSTCHPGTATISAGVSLPCLSAALL--TG	94
Human	46	IGS	MRPLPPLITVSGNSLPIHRTSTCHPGTATISAGVSLPCLSAALL--LPG	94
<i>C. elegans</i>	97	D	TAFVDLIRWNPCVTAFFYYIAFFIIFTLNGEFGFDARQVN	146
Bovine	95	S	FESHLEIFKRSCLQPAITATAPALVTPPLHHTGCRHLDNDLQKGL	143
Human	95	N	EENYLELTFKRSCLQPAITATAPALVTPPLHHTGCRHLDNDLQKGL	143
<i>C. elegans</i>	147	N	VGVIKCYL--SGSAI--ALAIVFNSCQNKSNKTA	182
Bovine	144	T	SLHNSGA--LVL--VGLAAH-----	169
Human	144	K	POLYOSGV--LVL--VGLAAH-----	169

Figure 3 Comparison of the predicted amino-acid sequences of *C. elegans* *cyt-1* (*mev-1*), bovine and human cytochrome *b*₅₆₀. Boxes indicate positions at which residues are identical. The sequence of human cytochrome *b*₅₆₀ was reconstructed from incomplete sequences in the GenBank database: accession numbers R87065, H83619 and N34060 for cDNA, and AC00406 for genomic DNA.

How does the *mev-1* mutation exert its effects on mitochondria and, ultimately, the nematode? This may relate to the reduction of coenzyme Q, because the principal mitochondrial source of free radicals is the interaction of the Q semiquinone with molecular oxygen, which occurs primarily at complex III to produce superoxide^{15,16}. In complex III, the function of cytochrome *b* is to remove the semiquinone by effectively acting as a Q dismutase in the protonmotive Q cycle; treatment with antimycin blocks this dismutase activity of complex III (ref. 16). The cytochrome *b*₅₆₀ in complex II may also have a similar function in stabilizing or dismuting the Q semiquinone: the *mev-1* mutation may abolish that function and thus allow semiquinone to build up. Its reaction with oxygen would result in excessive superoxide production, which would lead to the dual phenotypes of oxygen hypersensitivity and premature ageing.

Defective mitochondrial respiratory enzymes cause myopathy and neurological diseases in humans, with pathologies that include precocious ageing^{17,18}. Compared with defects in the other four complexes, abnormalities in complex II are rare and clinical presentation varies among individuals^{19–22}. In some cases, as in Leigh's syndrome²¹, a mutation occurs in SDH itself, but the basis of the other complex II human diseases is still not clear.

To our knowledge, the *mev-1* allele constitutes the first reported mutation of the SDH cytochrome *b* subunit in animals. It should provide additional insight into the molecular mechanisms of some of the mitochondrially related human myopathies, neurodegenerative diseases and premature ageing.

Methods

Measurements of survival and lifespan. Transgenic animals containing a 5.6-kb fragment (*kn1:knIs2*) were obtained from both oxygen-resistant and roller phenotypes by co-injection of the *rol-6* gene with a dominant mutation (*pRF4*). Integration of extrachromosomal DNA arrays into a chromosome was achieved by UV irradiation (254 nm, 300 J m⁻²) to parent strains (S. Mitani, personal communication). Survival: newly hatched first-stage (L1) larvae were exposed to various oxygen concentrations in an airtight plastic chamber. 4–6 days later, survival was determined by counting the number of fourth-stage (L4) larvae and adults. Lifespan: L1 larvae were grown under atmospheric oxygen at 25 °C until they reached adulthood. The lifespan of individual adult hermaphrodites was then determined under 60% oxygen at the same temperature.

RT-PCR amplification of *cyt-1* cDNA. Total RNA was prepared using standard techniques²³. RNAs were reverse-transcribed using superscript II RNase H reverse transcriptase (Gibco BRL). PCR (in 1.5 mM MgCl₂; annealing temperature, 60 °C) was performed for exons 1 and 2 using the following primers (forward, exon 1, 5'-TTC CAA CTG CGA TTC TGT GC-3'; reverse, exon 2, 5'-GCG GTG AAA TCG AAC GGC A-3').

Preparation of membrane fractions. Wild-type or *mev-1* *C. elegans* were collected by centrifugation at 1,500g for 10 min. Animals were washed with PBS three times and once with 0.15 M NaCl containing 20 mM HEPES–NaOH

buffer (buffer A), pH 7.4. The pellet was resuspended in two volumes of buffer A containing one proteinase-inhibitor cocktail tablet (Boehringer Mannheim) in 50 ml, and cells were disrupted by a Teflon homogenizer. This lysate was centrifuged at 1,500g for 10 min at 4 °C. The supernatant was then spun at 100,000g for 30 min. The pellet containing submitochondrial particles was resuspended in buffer A.

Assays for complex II and SDH (E.C.1.3.99.1) activity. Complex II activity was assayed by malonate-sensitive succinate–cytochrome *c* reductase activity²⁴ at 37 °C in 1.0 ml 0.1 M Tris–SO₄ buffer, pH 7.4, containing 1.6 mg cytochrome *c*, 1 mM sodium cyanide, 20 mM sodium succinate. The reference cuvette contained 20 µl of 20% sodium malonate solution. SDH activity was assayed as phenazine methosulphate (PMS)-mediated reduction of cytochrome *c*²⁵ in 1.0 ml 0.1 M Tris–SO₄ buffer, pH 7.4, containing 0.165 mg PMS, 0.8 mg *c*, 1 mM sodium cyanide, 20 mM sodium succinate, against a reference of 20 µl of 20% sodium malonate. In both assays, cytochrome *c* reduction was measured by the increase in absorbance at 550 nm.

Received 5 May; accepted 18 June 1998.

- Harman, D. Free radical theory of aging: effect of free radical reaction inhibitors on the mortality rate of male LAF mice. *J. Geront.* **23**, 476–482 (1968).
- Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of ageing: role of oxidative damage and environmental stresses. *Nature Genet.* **13**, 25–34 (1996).
- Larsen, P. L. Aging and resistance to oxidative damage in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 8905–8909 (1993).
- Agarwal, S. & Sohal, R. S. DNA oxidative damage and life expectancy in houseflies. *Proc. Natl Acad. Sci. USA* **91**, 12332–12335 (1994).
- McCord, J. M. & Fridovich, I. Superoxide dismutase: An enzymic function for erythrocuprein (hemocuprein). *J. Biol. Chem.* **244**, 6049–6055 (1969).
- Ishii, N. *et al.* A methyl viologen-sensitive mutant of the nematode *Caenorhabditis elegans*. *Mut. Res.* **237**, 165–171 (1990).
- Honda, S., Ishii, N., Suzuki, K. & Matsuo, M. Oxygen-dependent perturbation of life span and aging rate in the nematode. *J. Geront.* **48**, B57–B61 (1993).
- Hosokawa, H. *et al.* Rapid accumulation of fluorescent material with aging in an oxygen-sensitive mutant *mev-1* of *Caenorhabditis elegans*. *Mech. Ageing Dev.* **74**, 161–170 (1994).
- Adachi, H., Fujiwara, Y. & Ishii, N. Effects of oxygen on protein carbonyl and aging in *Caenorhabditis elegans* mutants with long (age-1) and short (mev-1) life spans. *J. Geront.* **53A**, B240–B244 (1998).
- Fire, A. Integrative transformation of *Caenorhabditis elegans*. *EMBO J.* **5**, 2673–2680 (1986).
- Cochran, B., Capaldi, R. A. & Ackrell, B. A. C. The cDNA sequence of beef heart C_{II}-3, a membrane-intrinsic subunit of succinate-ubiquinone oxidoreductase. *Biochim. Biophys. Acta* **1188**, 162–166 (1994).
- Hengartner, M. O. & Horvitz, H. R. *C. elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene *bcl-2*. *Cell* **76**, 665–676 (1994).
- Yu, L., Xu, J.-X., Haley, P. E. & Yu, C.-A. Properties of bovine heart mitochondrial cytochrome *b*₅₆₀. *J. Biol. Chem.* **262**, 1137–1143 (1987).
- Friden, H. & Hederstedt, L. Role of His residues in *Bacillus subtilis* cytochrome *b*₅₅₈ for haem binding and assembly of succinate:quinone oxidoreductase (complex II). *Mol. Microbiol.* **4**, 1045–1056 (1990).
- Cadenas, E. & Boveris, A. Enhancement of hydrogen peroxide formation by protophores and ionophores in antimycin-supplemented mitochondria. *Biochem. J.* **188**, 31–37 (1980).
- Turrens, J. F., Alexandre, A. & Lehninger, A. Ubisemiquinone is the electron donor for superoxide formation by complex III of heart mitochondria. *Arch. Biochem. Biophys.* **237**, 408–414 (1985).
- Wallace, D. C. Diseases of the mitochondrial DNA. *Annu. Rev. Biochem.* **61**, 1175–1212 (1992).
- Munnich, A. *et al.* Clinical aspects of mitochondrial disorders. *J. Inher. Metab. Dis.* **15**, 448–455 (1992).
- Riggs, J. E. *et al.* Mitochondrial encephalomyopathy with decreased succinate-cytochrome *c* reductase activity. *Neurology* **34**, 48–53 (1984).
- Martin, J. J. *et al.* Defect in succinate oxidation by isolated muscle mitochondria in a patient with symmetrical lesions in the basal ganglia. *J. Neurol. Sci.* **84**, 189–200 (1988).
- Bourgeron, T. *et al.* Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nature Genet.* **11**, 144–149 (1995).
- Taylor, R. W. *et al.* Deficiency of complex II of the mitochondrial respiratory chain in late-onset optic atrophy and ataxia. *Ann. Neurol.* **39**, 224–232 (1996).
- Chomczynski, P. & Sacchi, N. Single-step method for RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Analyt. Biochem.* **162**, 156–159 (1987).

24. Robinson, K. M. & Lemire, B. D. Flavinylation of succinate: ubiquinone oxidoreductase from *Saccharomyces cerevisiae*. *Meth. Enzymol.* **260**, 34–51 (1995).
25. Ackrell, B. A. C., Kearney, E. B. & Singer, T. P. Mammalian succinate dehydrogenase. *Meth. Enzymol.* **53**, 466–483 (1978).

Acknowledgements. We thank M. Hengartner for strains and for suggestions that facilitated the mapping of *mev-1*. The wild-type strain was from the *C. elegans* Genetics Center, which is supported by the National Center for Research Resources (NCRR). This work was supported by Tokai University School of Medicine Research Project and by a Grant in Aid for Aging Research from the Ministry of Human and Welfare, Japan, and for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan.

Correspondence and requests for materials should be addressed to N.I. (e-mail: nishii@is.icc.u-tokai.ac.jp).

Protein kinase C regulates the nuclear localization of diacylglycerol kinase- ζ

Matthew K. Topham^{*†‡}, Michaeline Bunting^{*†‡},
Guy A. Zimmerman[‡], Thomas M. McIntyre[‡],
Perry J. Blackshear[§] & Stephen M. Prescott^{*†‡}

^{*} The Huntsman Cancer Institute, [†] Eccles Program in Human Molecular Biology & Genetics, and [‡] Department of Internal Medicine, University of Utah, Salt Lake City, Utah 84112, USA

[§] National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA

Diacylglycerol kinases (DGKs) terminate signalling from diacylglycerol by converting it to phosphatidic acid^{1–8}. Diacylglycerol regulates cell growth and differentiation, and its transient accumulation in the nucleus may be particularly important in this regulation^{9,10}. Here we show that a fraction of DGK- ζ is found in the nucleus, where it regulates the amount of nuclear diacylglycerol. Reducing nuclear diacylglycerol levels by conditional expression of DGK- ζ attenuates cell growth. The nuclear-localization signal of DGK- ζ is located in a region that is homologous to the phosphorylation-site domain of the MARCKS protein. This is, to our knowledge, the first evidence that this domain, which is a major target for protein kinase C, can localize a protein to the nucleus. Two isoforms of protein kinase C, but not others, regulate the localization of DGK- ζ . Our results define a cycle in which diacylglycerol activates protein kinase C, which then regulates the metabolism of diacylglycerol by alternating the intracellular location of DGK- ζ . This may be a general mechanism to control mitogenic signals that depend on nuclear diacylglycerol.

Growth factors and other stimuli cause a transient increase in levels of cellular diacylglycerol (DAG), which activates protein kinase C (PKC). Activated PKC regulates growth and differentiation, and dysregulated activity is mitogenic. One route for terminating DAG signalling is its phosphorylation by diacylglycerol kinases^{1–8}, producing phosphatidic acid. The cycle of phospholipase activation, a rise in DAG levels, PKC activation and then conversion of DAG to phosphatidic acid has been studied most extensively at the plasma membrane. However, there is a parallel nuclear cycle, which may regulate the cell cycle^{9,10}. If so, nuclear DGK could regulate cellular proliferation. DGK- ζ has a motif similar to the phosphorylation-site domain (PSD) of the MARCKS (myristoylated alanine-rich C-kinase substrate) protein. The PSD in MARCKS is rich in basic amino acids and is phosphorylated by PKC^{11,12}. As this motif in DGK- ζ is similar to some nuclear-localization signals (NLSs), we proposed that DGK- ζ is a nuclear protein. We found 15–30% of total cellular DGK- ζ in the nucleus both of COS-7 cells transfected with DGK- ζ complementary DNA (Fig. 1a) and of A172 cells, a glioblastoma-derived line, which express native DGK- ζ (Fig. 2c). There was a similar distribution of DGK activity (not shown) in both cell types.

To determine the nuclear-targeting region of DGK- ζ , we con-

structed a cDNA lacking the MARCKS PSD homology domain, transfected COS-7 cells, and determined the location of DGK- ζ protein. Unlike the wild-type protein (Fig. 1b), DGK- ζ lacking the MARCKS PSD was almost entirely extranuclear (Fig. 1c). Deleting either the ankyrin repeats or the amino-terminal 120 amino acids did not alter the nuclear localization (not shown). Immunoblotting (as in Fig. 1a) showed that the PSD deletion resulted in a >70% decrease in nuclear protein compared with wild-type DGK- ζ (not shown). We then fused the deleted sequence to a green fluorescent protein (GFP) and assessed its localization. The PSD-fusion protein was found in the nucleus (Fig. 1e) whereas GFP fusions with other regions of DGK- ζ were predominantly extranuclear (not shown); GFP alone was uniformly distributed in the cell (Fig. 1d). Thus, a 104-amino-acid fragment containing the MARCKS domain (I245–I348) is both necessary and sufficient to direct DGK- ζ to the nucleus.

The PSD in MARCKS is phosphorylated by PKC, which suggested that DGK- ζ may also be a substrate for PKC, and that this could regulate its intracellular location. We incubated a peptide containing the wild-type sequence of the DGK- ζ PSD (K256 ASKKKKRASFKRKSSK K273), or the same peptide in which serine 10 (italic) was changed to alanine (A/S peptide), with an active fragment of PKC (PKM). Both peptides were phosphorylated, but phosphorylation of DGK- ζ A/S proceeded at only one-sixth the rate of phosphorylation of the wild-type peptide (not shown). Thus, the substituted serine (S265) is the primary site of PKC-dependent phosphorylation, as predicted by homology to MARCKS and MARCKS-related protein (MRP). The wild-type peptide had a K_m value of about 500 nM, which compares favourably with the values for the MARCKS and MRP peptides (20 and 173 nM, respectively)^{13,14}. In contrast to the MARCKS and MRP peptides, which are phosphorylated with pronounced positive cooperativity, phosphorylation of the DGK- ζ peptide showed conventional

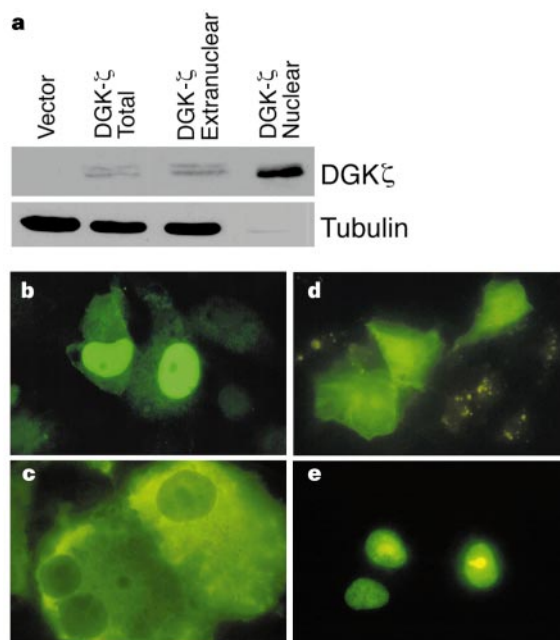


Figure 1 DGK- ζ is a nuclear protein and its MARCKS PSD is necessary and sufficient for nuclear targeting. **a**, Western blot of COS-7 cells transfected either with control vector or with DGK- ζ cDNA and then fractionated into nuclear and extranuclear components (15 μ g protein per lane). The blot was probed with either anti-DGK- ζ (upper panel) or anti-tubulin (lower panel) antibodies. **b**, **c**, Immunocytochemistry detecting DGK- ζ in COS-7 cells transfected either with wild-type DGK- ζ (**b**) or with a construct the MARCKS PSD was deleted (**c**). **d**, **e**, Regions of DGK- ζ were individually fused to the terminus of GFP and transfected into COS-7 cells. **d**, GFP alone was transfected. **e**, GFP-fusion protein was transfected.

Michaelis–Menten kinetics, with a Hill constant of 0.7 (data not shown). The DGK A/S peptide possibly had biphasic kinetics, indicating that it was phosphorylated at one or two other, lower affinity, sites. As predicted from studies of the MARCKS tetra-ala peptide¹⁴, the DGK- ζ A/S peptide inhibited PKM when histone H1S was the substrate ($K_i \sim 800$ nM).

To test whether this phosphorylation regulates nuclear localization of DGK- ζ , we expressed in COS-7 cells the full-length DGK- ζ with PKC- α or PKC- γ , which both phosphorylate MARCKS¹⁵ and are expressed in A172 cells¹⁶. PKC- α (Fig. 2a) and PKC- γ (not shown) both reduced the fraction of DGK- ζ in the nucleus by 30 to 50%. We further tested whether the nuclear localization of DGK- ζ was regulated in the same way in A172 cells by downregulating these PKC isozymes by prolonged treatment with phorbol 12-myristate 13-acetate (PMA), and then determining the location of DGK- ζ . This procedure resulted in complete loss of these PKC isozymes (Fig. 2b), and the localization of DGK- ζ to the nucleus was

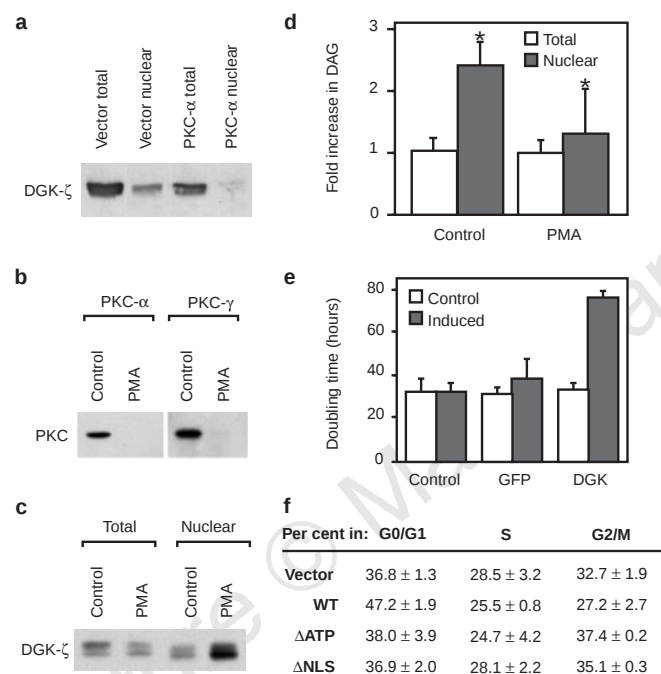


Figure 2 PKC controls nuclear localization of DGK- ζ , and overexpression of DGK- ζ regulates nuclear DAG and the cell cycle. **a**, Immunoblot (for DGK- ζ) of total cellular homogenates or nuclear fractions (10 μ g protein per lane) from COS-7 cells into which DGK- ζ had been transfected with control vector or with PKC- α . The cells were treated with PMA for 30 min before collecting. **b**, We treated A172 cells with 60 ng ml⁻¹ PMA or vehicle for 24 h, and then immunoblotted total cellular homogenates (10 μ g per lane) using anti-PKC- α or anti-PKC- γ antibodies. **c**, Immunoblot for DGK- ζ in A172 cells treated with PMA as in **b** (10 μ g per lane total and 20 μ g per lane nuclear). **d**, A172 cells were treated with PMA as in **b** and then with EGF (40 ng per ml) for 10 min. Nuclear and total cellular lipids were extracted and DAG/phosphate levels determined. Values are fold-increase over basal levels ($n = 4$). Asterisk indicates significance ($P < 0.05$). **e**, Doubling times of 293 cells stably expressing either vector, GFP, or DGK- ζ with or without inducing agent (1 μ M muristerone A). The values presented are the averages from two experiments. **f**, COS-7 cells were transfected with either control vector, wild-type DGK- ζ (WT), kinase-dead DGK- ζ (Δ ATP), or a MARCKS PSD mutant (Δ NLS) and then cell-cycle analysis was performed by flow cytometry. Mean values from two experiments are shown.

enhanced (Fig. 2c), as was nuclear DGK activity (not shown).

Growth factors and other stimuli cause turnover of phosphatidylinositol-4,5-bisphosphate and a rapid increase in DAG levels in cell nuclei^{17–20}, which can regulate growth and differentiation. As PKC regulates the amount of DGK- ζ in the nucleus, we next asked whether this DGK activity could attenuate the accumulation of nuclear DAG and modulate cell growth. In A172 cells treated with epidermal growth factor (EGF), nuclear DAG levels rose two- to threefold above baseline by 10 minutes after treatment, whereas total cellular levels were unchanged (not shown). To determine the function of DGK activity in the nucleus, we compared nuclear DAG levels in control A172 cells with levels in cells with high nuclear DGK- ζ levels (cells exposed to PMA; Fig. 2c). When the A172 cells were treated with EGF for 10 minutes, nuclear DAG in cells not pretreated with PMA reached 241% ($\pm 37\%$) of that in unstimulated cells, but nuclear DAG in cells that had been pretreated with PMA (that is, cells with high nuclear DGK- ζ levels) reached only

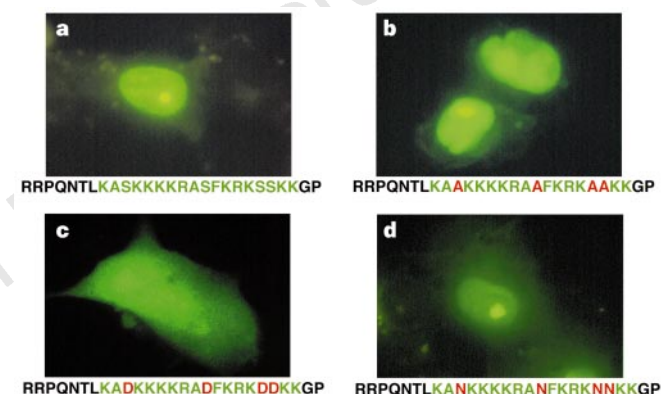


Figure 3 Introduction of negative charge into the MARCKS PSD excludes it from the nucleus. **a–d**, COS-7 cells were transfected with GFP fusions containing either the wild-type NLS (**a**) or constructs in which serines were altered to alanines (**b**), aspartates (**c**) or asparagines (**d**). The cells were then examined by fluorescence microscopy. The sequences are shown with altered amino acids in red and the MARCKS PSD in green.

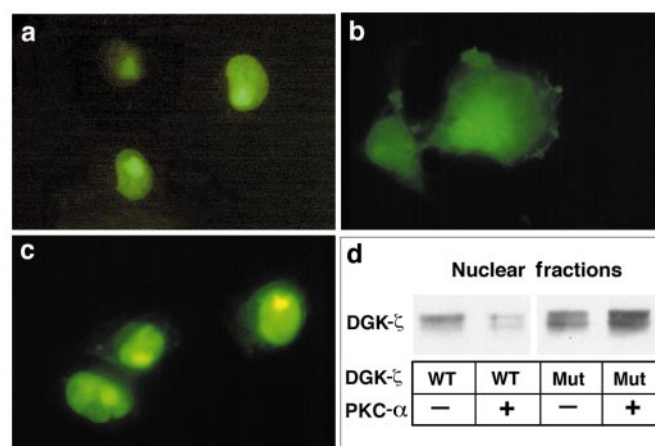


Figure 4 Phosphorylation by PKC- α causes nuclear exclusion of DGK- ζ . **a, b**, A GFP construct containing wild-type DGK- ζ NLS (Fig. 3a) was transfected into COS-7 cells together with control vector (**a**) or PKC- α (**b**). **c**, A GFP-NLS mutant with all serines changed to alanines (Fig. 3b) was transfected together with PKC- α . **d**, Western blot, using an anti-DGK- ζ antibody, of nuclear fractions (10 μ g protein per lane) into which control vector or PKC- α was transfected together with full-length wild-type (WT) DGK- ζ , or a full-length mutant DGK- ζ construct with all serines of the MARCKS PSD changed to asparagines (Mut). Controls (not shown) showed similar expression levels of total DGK- ζ .

130% ($\pm 72\%$) of unstimulated cells (Fig. 2d, $P < 0.05$). This supported the hypothesis that DGK- ζ in the nucleus regulates the concentration of DAG there. To verify these observations, we generated a stable cell line with inducible overexpression of DGK- ζ . When the cells were induced, nuclear DAG levels fell by 38%; in control cells nuclear DAG levels were unchanged with induction (not shown). As DAG levels may be an important determinant of cell growth⁹, we tested the effect of DGK- ζ on doubling time; overexpression increased doubling time from 33.5 (± 3.5) to 77.0 (± 3.0) hours (Fig. 2e, $P < 0.05$). Similar results were obtained with another DGK- ζ -overexpressing clone, but not with cell lines stably transfected either with vector without insert (no change with induction) or with an inducible GFP (1.2-fold increase in doubling time with induction).

As a separate test of whether nuclear targeting, and DGK activity, were essential for the effects of DGK- ζ on the cell cycle, we transfected cDNAs encoding wild-type DGK- ζ , a kinase-dead mutant (Δ ATP, G355 $\rightarrow$ D355), or a mutant that did not localize to the nucleus (Δ NLS; see below) along with GFP as a reporter. The cells were collected at 48 hours, stained with propidium iodide, and then sorted by flow cytometry to identify transfected cells (GFP-positive). The cell-cycle profile of the transfected cells show that overexpression of wild-type DGK- ζ , but not of the kinase-dead or NLS mutants, induced an accumulation of the cells in the G0/G1 phases of the cell cycle (Fig. 2f). In another experiment, we synchronized the cells with nocodazole and saw a qualitatively similar, but more marked, accumulation in cells in G0/G1 (not shown). We conclude that DGK- ζ regulates cell growth and that both its enzymatic activity and its nuclear localization are essential.

We then further tested whether the phosphorylation of DGK- ζ by PKC, and subsequent intracellular relocation of DGK- ζ , is mediated at the MARCKS domain. We transfected cells with a plasmid encoding GFP fused to a 27-amino-acid sequence that contained the MARCKS PSD of DGK- ζ . Like the 104-amino-acid fragment (see above), this region directed GFP to the nucleus (see below). Changing the basic amino acids at either end of the MARCKS-homology domain abolished the nuclear localization of GFP. As the GFP-fusion proteins could have passively diffused into the nucleus, the nuclear localization may have resulted from a charge interaction there. To address this, we mutated to alanine the basic amino acids of the MARCKS PSD of full-length DGK- ζ (Δ NLS, A256 ASAAAAAASFAAASSA A273), transfected into COS-7 cells, and immunoblotted subcellular fractions. These K/R $\rightarrow$ A mutations resulted in $>90\%$ reduction in levels of nuclear protein (not shown). Thus, the MARCKS-homology domain of DGK- ζ is its predominant nuclear-localization signal.

To verify that phosphorylation of this region relocates DGK- ζ , we transfected GFP constructs containing either the native 27-amino-acid sequence, or one in which the serines had been changed to aspartates, which mimics phosphorylation²¹. The S $\rightarrow$ D GFP fusion protein was extranuclear, whereas an uncharged, structurally similar construct (S $\rightarrow$ N) was nuclear, as was an S $\rightarrow$ A construct (Fig. 3). Thus, phosphorylation of the serine residues in the MARCKS-homology region of DGK- ζ regulates its intracellular location—the unphosphorylated form is in the nucleus whereas the phosphorylated form is extranuclear. As these constructs were not truly phosphorylated, we transfected cells with cDNAs encoding one of eight PKC isoforms (α , β I, β II, γ , δ , ϵ , μ or ζ) with GFP-reporter constructs containing wild-type DGK- ζ NLS or the same sequence with serines changed to alanines. In agreement with Fig. 2a, when the native NLS-GFP fusion was with PKC- α or PKC- γ , it was redistributed from the nucleus to the cytoplasm (Fig. 4a–c). In contrast, the GFP construct in which the NLS serines had been replaced with alanines was nuclear, verifying that the phosphorylation occurred at the MARCKS domain. Expression of the native NLS-GFP with the other isoforms of human PKC did not influence the location (not shown). To ensure that the GFP-reporter constructs reflected the behaviour of the native protein, we expressed

PKC- α with a full-length construct of DGK- ζ in which the serines in the MARCKS PSD had been mutated to asparagines (K256 ANKKKKRANFKRKNNK K273) so that this region could no longer be phosphorylated. Unlike the wild-type protein, the amount of this construct in the nucleus was not affected by expression of PKC (Fig. 4d).

Lipid signalling pathways regulate cell growth and differentiation, and transient accumulation of DAG, particularly in the nucleus, is important for this regulation^{18,19,22,23}. Our results show that this signal is turned off by DGK- ζ . This requires DGK- ζ to be active and in the nucleus, a localization that is dynamically regulated by phosphorylation of the MARCKS-homology domain by some, but not all, isoforms of PKC. $\square$

Methods

cDNA constructs. We cloned DGK- ζ cDNA into pcDNA1/Amp or pcDNA3.1/Zeo plasmids (Invitrogen). To delete the MARCKS domain, this cDNA was digested with *Bam*HI and then religated. We constructed MARCKS-domain mutants of full-length DGK- ζ by incorporating a new *Eag*I site (guanine was altered to cytosine at nucleotide 919) and then ligating annealed oligonucleotides with desired mutations into the *Bam*HI/*Eag*I-digested cDNA. The Δ ATP mutant was made by a guanine-to-adenosine change at nucleotide 1152 using a Muta-Gene kit (Bio-Rad). Coding regions of DGK- ζ were fused to the carboxy terminus of pEGFP (Clontech) using the following restriction sites: *Eco* V/*Bam*HI (amino acids 2–245), *Bam*HI/*Bam*HI (amino acids 245–348), *Apa*I/*Xba*I (amino acids 333 to stop codon), and *Sac*I/*Xba*I (amino acids 685 to stop codon). To generate the GFP-NLS constructs, forward and reverse oligonucleotides were synthesized, annealed, and ligated into pEGFP vectors.

Transfections. Transfections were performed and cells collected as described⁷. In the PKC co-transfection experiments, 500 ng of PKC- α , β I, β II, γ , δ , ϵ , μ or ζ (in pcDNA3/Amp from J. Metherall) or control vector were combined with 500 ng of either full-length DGK- ζ or a GFP construct. In the GFP experiments with PKC, the transfection medium was removed at 5 h and replaced with 60 ng ml⁻¹ PMA or ethanol vehicle in DMEM medium for another 18 h. In the experiments with full-length DGK- ζ and PKC constructs, the transfection medium was removed at 5 h and replaced with DMEM for another 18 h. At that point, PMA was added for 30 min and the cells were collected.

Immunoblotting and immunocytochemistry. An anti-peptide antibody was generated in rabbits⁷ and was affinity-purified. Immunoblots were performed as described⁷. Anti-PKC antibodies (Transduction Laboratories) were used at the recommended dilutions. Bands on some blots were quantified by densitometry using Kodak 1D image-analysis software. Immunocytochemistry was performed as described²⁴. Cells transfected with GFP alone were washed and fixed as above. A172 cells (from ATCC) were stained as above except that the incubation times with anti-DGK- ζ and goat anti-rabbit antibodies were 90 and 45 min at room temperature.

Stable transfects. Cell lines that could be induced to overexpress DGK- ζ or GFP were made using the ecdysone system (Invitrogen). Full-length DGK- ζ cDNA was cloned into pIND(SP1). This construct or pIND(SP1) without insert was transfected into EcR 293 cells (Invitrogen), which were then selected in 400 μ g ml⁻¹ G418 (Life Technologies) and 400 μ g ml⁻¹ zeocin (Invitrogen). After induction with muristerone A, colonies were screened for DGK- ζ by immunocytochemistry, or for GFP by fluorescent phase-contrast microscopy. Appropriate colonies were replated and selected twice more until a homogenous, inducible population was verified.

Nuclear isolation. COS-7 and A172 nuclear isolation was done as described²⁵, typically from two 35-mm wells. Immunoblots of cell homogenates were then carried out or DAG levels were determined. To assess the purity of these nuclear isolates, immunoblots for tubulin (anti-tubulin antibody Clontech) were performed or lipid phosphate levels were determined. Nuclear lipid phosphate was typically 2–5% of total lipid phosphate, which concurred with previous reports²⁶.

DAG assay. Stable cell lines treated with inducing agent (1 μ M muristerone A) or control vehicle, or A172 cells treated with PMA or control vehicle, were grown to 80% confluence. For total cellular DAG, cells from one 35-mm well were collected in 2 ml methanol. Nuclei were collected using two 35-mm wells and the final nuclear pellet was brought into 2 ml methanol. Lipid extractions were done as described²⁷. DAG levels were determined using the DAG assay

system (Amersham), and phosphate levels also were determined from the lipid extracts²⁸.

Synthesis, purification and phosphorylation of peptides. Peptides were synthesized and phosphorylated by mixed rat brain active PKC isozymes (PKM) for 10 min at 30 °C (refs 13, 14). When the inhibitory effects of the DGK A/S peptide (0–100 µM) were assayed, histone H1S (final concentration 20 µg ml⁻¹) was used as a substrate for PKM. This reaction was linear for at least 40 min; we used 10 and 15 min time points to evaluate inhibitory effects of the peptide²⁹.

Received 27 April; accepted 28 May 1998.

- Goto, K. & Kondo, H. Molecular cloning and expression of a 90-kDa diacylglycerol kinase that predominantly localizes in neurons. *Proc. Natl Acad. Sci. USA* **90**, 7598–7602 (1993).
- Kai, M., Sakane, F., Imai, S.-I., Wada, I. & Kanoh, M. Molecular cloning of a diacylglycerol kinase isozyme predominantly expressed in human retina with a truncated and inactive enzyme expression in most other human cells. *J. Biol. Chem.* **269**, 18492–18498 (1994).
- Sakane, F., Yamada, K., Kanoh, H., Yokoyama, C. & Tanabe, T. Porcine diacylglycerol kinase sequence has zinc finger and E-F hand motifs. *Nature* **344**, 345–348 (1990).
- Houssa, B. *et al.* Cloning of a novel human diacylglycerol kinase (DGK0) containing three cysteine-rich domains, a proline-rich region and a pleckstrin homology domain with overlapping ras-associating domain. *J. Biol. Chem.* **272**, 10422–10428 (1997).
- Klauck, T. M., Xu, X., Mousseau, B. & Jaken, S. Cloning and characterization of a glucocorticoid-induced diacylglycerol kinase. *J. Biol. Chem.* **271**, 19871–19878 (1996).
- Sakane, F., Imai, S. I., Kai, M., Wada, I. & Kanoh, H. Molecular cloning of a novel diacylglycerol kinase isozyme with a pleckstrin homology domain and a C-terminal tail similar to those of the EPH family of protein tyrosine kinases. *J. Biol. Chem.* **271**, 8394–8401 (1996).
- Bunting, M., Tang, W., Zimmerman, G. A., McIntyre, T. M. & Prescott, S. M. Molecular cloning and characterization of a novel human diacylglycerol kinase. *J. Biol. Chem.* **271**, 10230–10236 (1996).
- Tang, W., Bunting, M., Zimmerman, G. A., McIntyre, T. M. & Prescott, S. M. Molecular cloning of a novel human diacylglycerol kinase highly selective for arachidonate-containing substrates. *J. Biol. Chem.* **271**, 10237–10241 (1996).
- Jackowski, S. Cell cycle regulation of membrane phospholipid metabolism. *J. Biol. Chem.* **271**, 20219–20222 (1996).
- Olson, E. N., Burgess, R. & Staudinger, J. Protein kinase C as a transducer of nuclear signals. *Cell Growth Differ.* **4**, 699–705 (1993).
- Aderem, A. The MARCKS brothers: a family of protein kinase C substrates. *Cell* **71**, 713–716 (1992).
- Blackshear, P. J. The MARCKS family of cellular protein kinase C substrates. *J. Biol. Chem.* **268**, 1501–1504 (1993).
- Verghese, G. M. *et al.* Protein kinase C-mediated phosphorylation and calmodulin binding of recombinant MARCKS and MARCKS-related protein (MRP). *J. Biol. Chem.* **269**, 9361–9367 (1994).
- Gaff, J. M., Rajan, R. R., Randall, R. R., Nairn, A. C. & Blackshear, P. J. Protein kinase C substrate and inhibitor characteristics of peptides from the myristoylated alanine-rich C kinase substrate (MARCKS) protein phosphorylation site domain. *J. Biol. Chem.* **266**, 14390–14398 (1991).
- Herget, T., Oehlein, S. A., Pappin, D. J. C., Rozengurt, E. & Parker, P. J. The myristoylated alanine-rich C-kinase substrate (MARCKS) is sequentially phosphorylated by conventional, novel and atypical isoforms of protein kinase C. *Eur. J. Biochem.* **233**, 448–457 (1995).
- Xiao, H., Goldthwait, D. A. & Mapstone, T. The identification of four protein kinase C isoforms in human glioblastoma cell lines: PKC alpha, gamma, epsilon, and zeta. *J. Neurosurg.* **81**, 734–740 (1994).
- Divecha, N., Banfic, H. & Irvine, R. F. Inositides and the nucleus and inositides in the nucleus. *Cell* **74**, 405–407 (1993).
- Divecha, N., Banfic, H. & Irvine, R. F. The polyphosphoinositide cycle exists in the nuclei of Swiss 3T3 cells under the control of a receptor (for IGF-1) in the plasma membrane, and stimulation of the cycle increases nuclear diacylglycerol and apparently induces translocation of protein kinase C to the nucleus. *EMBO J.* **10**, 3207–3214 (1991).
- Banfic, H., Zizak, M., Divecha, N. & Irvine, R. F. Nuclear diacylglycerol is increased during cell proliferation *in vivo*. *Biochem. J.* **290**, 633–636 (1993).
- Leach, K. L. *et al.* α-Thrombin stimulates nuclear diglyceride levels and differential nuclear localization of protein kinase C isozymes in IIC9 cells. *J. Biol. Chem.* **267**, 21816–21822 (1992).
- Swierczynski, S. L. & Blackshear, P. J. Myristoylation-dependent and electrostatic interactions exert independent effects on the membrane association of the myristoylated alanine-rich protein kinase C substrate protein in intact cells. *J. Biol. Chem.* **271**, 23424–23430 (1996).
- Sun, B., Murray, N. R. & Fields, A. P. A role for nuclear phosphatidylinositol-specific phospholipase C in the G2/M phase transition. *J. Biol. Chem.* **272**, 26313–26317 (1997).
- Divecha, N., Rhee, S.-G., Letcher, A. J. & Irvine, R. F. Phosphoinositide signalling in rat liver nuclei: phosphoinositidase C isoform β1 is specifically, but not predominantly, located in the nucleus. *Biochem. J.* **289**, 617–620 (1993).
- Ding, L. *et al.* Alternative splicing of the human diacylglycerol kinase zeta gene in muscle. *Proc. Natl Acad. Sci. USA* **94**, 5519–5524 (1997).
- Payrastra, B. *et al.* A differential location of phosphoinositide kinases, diacylglycerol kinase, and phospholipase C in the nuclear matrix. *J. Biol. Chem.* **267**, 5078–5084 (1992).
- York, J. & Majerus, P. W. Nuclear phosphatidylinositols decrease during S-phase of the cell cycle in HeLa cells. *J. Biol. Chem.* **269**, 7847–7850 (1994).
- Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
- Whately, R. E. *et al.* Growth-dependent changes in arachidonic acid release from endothelial cells are mediated by protein kinase C and changes in diacylglycerol. *J. Biol. Chem.* **268**, 16130–16138 (1993).
- Halsey, D. L., Girard, P. R., Kuo, J. F. & Blackshear, P. J. Protein kinase C in fibroblasts. Characteristics of its membrane association during growth and after exposure to phorbol esters and other mitogens. *J. Biol. Chem.* **262**, 2234–2243 (1987).

Acknowledgements. We thank E. Kennington, M. Flynn, J. White and H. Rust for technical assistance, M. McAdams for peptide synthesis, and V. Bennett and A. Thorburn for discussions. This work was supported by a grant from the National Cancer Institute. The core facilities at the University of Utah (DNA Sequencing and Peptide/DNA User Facility) were supported by grants from the National Cancer Institute. P.J.B. was an Investigator of the Howard Hughes Medical Institute when this work was performed, and M.K.T. is a Howard Hughes Medical Institute Physician Postdoctoral Fellow.

Correspondence and requests for materials should be addressed to S.M.P. (e-mail: Steve.Prescott@genetics.utah.edu).

DNA-dependent protein kinase acts upstream of p53 in response to DNA damage

Richard A. Woo*, Kevin G. McLure*, Susan P. Lees-Miller†, Derrick E. Rancourt‡ & Patrick W. K. Lee*

Departments of * Microbiology and Infectious Diseases, † Biological Sciences, and ‡ Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta T2N 4N1, Canada

The tumour suppressor p53 becomes activated as a transcription factor in response to DNA damage^{1–3}, but the mechanism for this activation is unclear. A good candidate for an upstream activator of p53 is the DNA-dependent protein kinase (DNA-PK) that depends on the presence of DNA breaks for its activity^{4–6}. Here we investigate the link between DNA damage and the activation of DNA-PK and of p53. To determine whether DNA-PK is an upstream mediator of the p53 DNA-damage response, we analysed a severe combined-immunodeficiency (SCID) mouse cell line, SCGR11 (refs 7, 8), and the human glioma cell line M059J (ref. 9). Both cell lines lack any detectable DNA-PK activity. We find that p53 is incapable of binding to DNA in the absence of DNA-PK, that DNA-PK is necessary but not sufficient for activation of

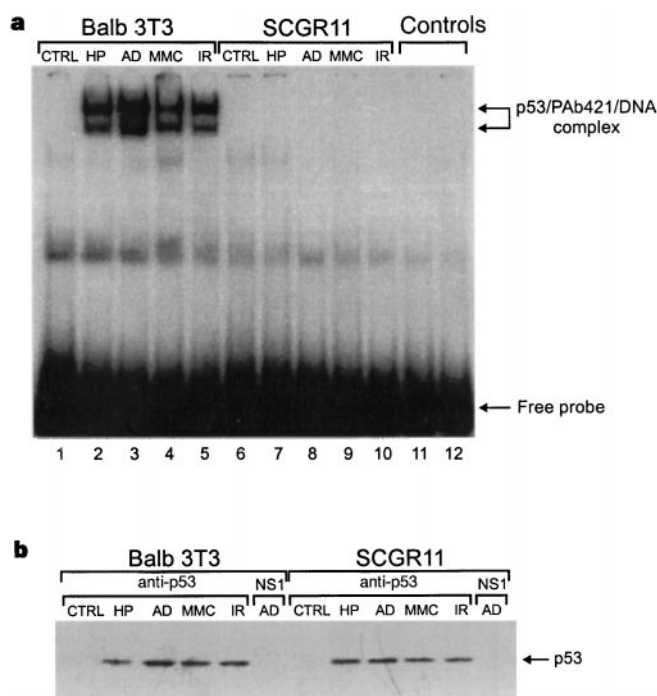


Figure 1 The effect of DNA damage on p53 DNA-binding activity in nuclear extracts from DNA-PK-deficient SCGR11 cells. **a**, BALB/c 3T3 (lanes 1–5) or SCGR11 (lanes 6–10) cells were exposed to different DNA-damaging agents (mock-treated, CTRL; hydrogen peroxide, HP; actinomycin D, AD; mitomycin C, MMC; γ-irradiation, IR). Nuclear extracts were then prepared and assayed for p53 DNA-binding by EMSA. All reactions contained pAb421 unless otherwise stated. 'Control' lanes are as follows: lane 11 is identical to lane 3, except pAb421 was not used; lane 12 contained the same nuclear extract as lane 3, except that a nonspecific DNA oligonucleotide probe (NB in ref. 23) was used in the place of the p53 consensus binding oligonucleotide probe. **b**, BALB/c 3T3 or SCGR11 cells were mock-treated (CTRL) or DNA-damaged as in **a** and then labelled with ³⁵S-methionine for 2 h. Nuclear extracts were immunoprecipitated with either anti-p53 monoclonal antibodies or control antibody (NS1).

p53 sequence-specific DNA binding, and that this activation occurs in response to DNA damage. Our results establish DNA-PK as a link between DNA damage and p53 activation, and reveal the existence of a mammalian DNA-damage-response pathway.

To determine whether DNA-PK is an upstream regulator of p53, we compared the mouse SCID cell line SCGR11, which contains no detectable DNA-PK activity⁸, with the wild-type BALB/c 3T3 cell line. The cells were exposed to different DNA-damaging agents (hydrogen peroxide, actinomycin D, mitomycin C, or γ -irradiation) and nuclear extracts were assayed for p53 sequence-specific DNA-binding by electrophoretic mobility shift assay (EMSA) (Fig. 1a). As expected, p53 from DNA-damaged BALB/c 3T3 cells was activated to bind to its consensus binding sequence (Fig. 1a: compare lane 1 with lanes 2, 3, 4 and 5). However, there was no p53 DNA binding in nuclear extracts of DNA-damaged SCGR11 cells (Fig. 1a, lanes 7–10), although p53 in these cells is wild-type and fully capable of binding to the consensus sequence under appropriate conditions (see below). Metabolic labelling with ³⁵S-methionine revealed that newly synthesized p53 accumulated to a comparable extent in both BALB/c 3T3 and SCGR11 cells after DNA damage (Fig. 1b). Immunoblot analysis also revealed that p53 expression was very high in untreated SCGR11 cells relative to the BALB/c 3T3 cells (data not shown). These observations therefore challenge the current idea that stabilization of p53 implies that it is activated upon DNA damage.

To test whether DNA-binding activity *in vitro* corresponds to activation of transcription *in vivo*, we transiently transfected both the BALB/c 3T3 and the SCGR11 cells with pG₁₃CAT, a chloram-

phenicol acetyl transferase (CAT) reporter construct controlled by a p53-responsive element¹⁰. In both BALB/c 3T3 and SCGR11 cells, there was minimal expression of the CAT reporter construct in untreated cells (Fig. 2a). Following exposure to different DNA-damaging agents, BALB/c 3T3 cells had high CAT activity, whereas the DNA-damaged SCGR11 cells were indistinguishable from the untreated control.

As activated p53 can transactivate the cyclin-dependent-kinase inhibitor p21 (CIP1)^{11,12}, we measured the relative expression of p21 messenger RNA in untreated and DNA-damaged cells. The results (Fig. 2b) show that p21 gene expression was upregulated in BALB/c 3T3 cells following treatment with DNA-damaging agents, but not in DNA-damaged SCGR11 cells relative to untreated SCGR11 cells. These results are consistent with those of the CAT assay. We conclude that DNA-PK is an upstream activator of p53 sequence-specific DNA-binding activity and therefore affects p53-dependent transactivation of downstream effectors, including p21.

It has been proposed that DNA-PK is not an upstream regulator of p53 because primary SCID lymphocytes and fibroblasts respond to ionizing radiation by inducing p53 expression and either arresting at the G1 phase of the cell cycle or undergoing apoptosis^{13–16}. However, the DNA-PK status of primary SCID cells has not been determined and has been defined in only two different SCID cell lines, including the SCGR11 cells used here. We found a low but detectable level of DNA-PK activity in mouse embryonic fibroblasts (MEFs) from SCID mice (Fig. 3a); this activity was about half of that found in MEFs from BALB/c mice from which the SCID mice were derived, and ~15% of that in BALB/c 3T3 cells. SCGR11 cells had

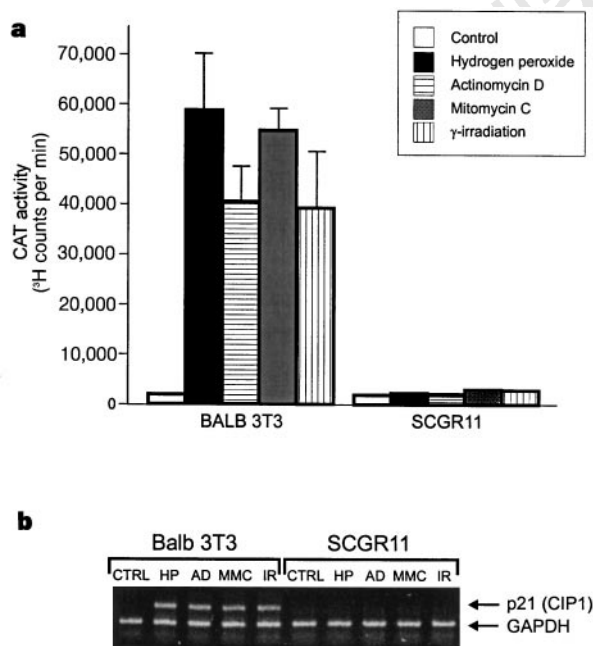


Figure 2 The effect of DNA damage on p53-dependent transactivation in the SCGR11 cells. **a**, BALB/c 3T3 and SCGR11 cells were transiently cotransfected with pG₁₃CAT and pCH110 (Pharmacia). Following transfection, cells were either mock-treated (CTRL) or DNA-damaged as in Fig. 1a. CAT activity is expressed as counts per minute of ³H-labelled acetylated chloramphenicol. Data are the average from three independent experiments. **b**, Relative p21 (CIP1) expression was determined for BALB/c 3T3 and SCGR11 cells following mock-treatment or DNA damage as in Fig. 1a. Total RNA was amplified by RT-PCR, followed by selective amplification of p21 (CIP1) and the constitutively expressed GAPDH, which served as a PCR and gel-loading control.

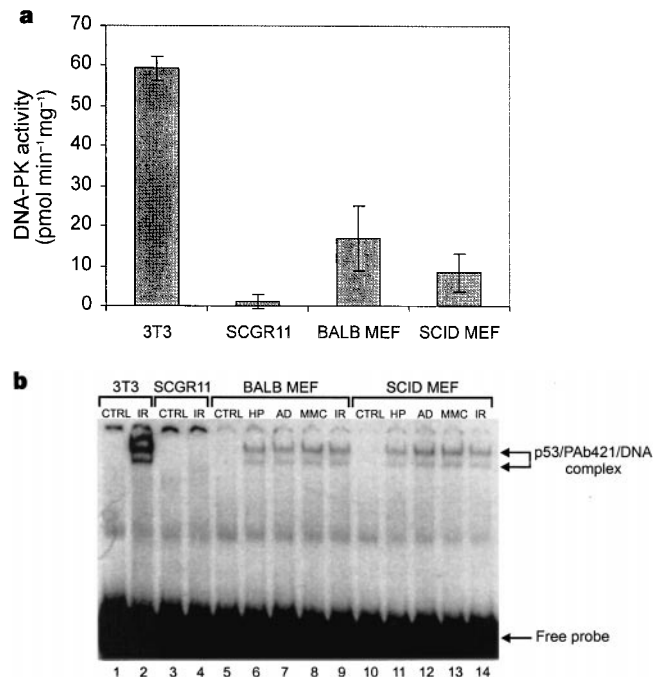


Figure 3 Comparing primary SCID cells and the SCGR11 SCID cell line in terms of DNA-PK activity and p53 DNA-binding activity. **a**, DNA-PK activity was determined for BALB/c and SCID MEFs as well as the BALB/c 3T3 and SCGR11 cell lines. DNA-PK activity is expressed as picomol ³²P-ATP transferred to the substrate peptide per min, per mg protein. Data are the average from three independent experiments. Analysis (using the Student's *t*-test) of the data sets for the SCGR11 cells and the SCID MEFs gave a 98% confidence interval that the DNA-PK activity is statistically different between the two cell types. **b**, Primary BALB/c MEFs (lanes 5–9) or SCID MEFs (lanes 10–14) were mock-treated (CTRL) or DNA-damaged with HP, AD, MMC or IR, as in Fig. 1a. Nuclear extracts were then prepared and assayed for p53 DNA-binding by EMSA. Nuclear extracts from mock-treated or γ -irradiated 3T3 and SCGR11 cells were also used for comparison (lanes 1–4).

no detectable DNA-PK activity⁸. The low level of DNA-PK activity in primary SCID cells was probably sufficient to activate p53 in the earlier studies^{13–16}. We confirmed that p53 in these cells is activatable when assaying for p53 DNA-binding activity in DNA-damaged SCID MEFs (Fig. 3b, lanes 11–14). These results therefore challenge the contention that primary SCID cells are devoid of DNA-PK activity, and may explain why Ku knockout mice¹⁷ and true DNA-PK_{CS} knockout mice¹⁸ have defects not found in SCID mice. There is a difference between primary SCID MEFs, which are not true DNA-PK_{CS} knockout cells, and established SCID cell lines such as SCGR11.

To verify this connection between DNA-PK and p53, we investigated another cell line that also lacks DNA-PK activity, the human glioma cell line M059J. Like SCGR11 cells, M059J cells lack any DNA-PK activity owing to defective expression of DNA-PK_{CS}⁹. The control cell line M059K was derived from a different section of the same glioma, but these cells have the intact DNA-PK holoenzyme (that is, DNA-PK_{CS} and Ku) and activity⁹. We note that both cell lines express p53 (as detected by western blotting), but in neither case can it bind DNA, nor does it accumulate after DNA damage (data not shown); these cells probably express a mutant form of p53 as p53 is mutated in most human cancers¹⁹. We then used these cells in a cell-free *in vitro* translation system, and found that cytoplasmic extracts from each cell line could translate exogenous wild-type p53 at comparable levels (data not shown). Following *in vitro* synthesis of p53 in the cytoplasmic extracts, nuclear extract from each glioma cell line (untreated or γ -irradiated) was added to test whether the translated p53 could be activated for DNA binding. Endogenous p53 (probably mutated) in these nuclear extracts was first immunodepleted to avoid interference with exogenous translated wild-type p53. As controls, nuclear extracts (from either untreated or γ -irradiated cells) from each glioma cell line, in combination with mock-translated cytoplasmic extract (without exogenous p53

mRNA), did not produce a DNA-mobility shift in the EMSA (Fig. 4, lanes 1–4). When nuclear extracts from untreated M059K or M059J cells were added to p53 synthesized *in vitro*, there was still no DNA binding (Fig. 4, lanes 5, 7, 9 and 11). However, when nuclear extract from γ -irradiated M059K cells was added to *in vitro* translated p53, there was a pronounced activation of p53 DNA-binding (Fig. 4, lanes 6 and 12) which did not occur when γ -irradiated nuclear extract from DNA-PK-deficient M059J cells was added (Fig. 4, lanes 8 and 10). This indicates that both DNA damage and the presence of DNA-PK are necessary for p53 activation and that this activation occurs through a post-translational mechanism. We confirmed this by adding γ -irradiated M059K nuclear extract to SCGR11 nuclear extract and showing that endogenous p53 in SCGR11 cells could be similarly activated (Fig. 4: compare lanes 13 and 14). Also, the results show that p53 in SCGR11 cells is wild-type.

To confirm that the kinase function of DNA-PK is involved in p53 activation, we used a substrate peptide of DNA-PK²⁰ as a competitive inhibitor of DNA-PK (Fig. 5a). When we added this substrate peptide to DNA-damaged nuclear extract from γ -irradiated M059K cells, p53 DNA-binding was blocked (Fig. 5a: compare lanes 2 and 3; no effect with control peptide, lane 4), demonstrating the importance of DNA-PK kinase function in p53 activation.

The most logical explanation for the dual requirement of DNA damage and DNA-PK for p53 activation would be that DNA-PK needs to be first activated by DNA breaks before it can activate p53, directly or indirectly. If the effect of DNA damage is strictly the generation of DNA breaks for DNA-PK activation, then DNA-PK holoenzyme alone should activate p53 in standard p53 mobility-shift assays in which DNA oligonucleotides (known to activate DNA-PK^{8,9,20}) are present. We found that p53 synthesized in a cytoplasmic extract of M059J cells was not activated by DNA-PK holoenzyme alone under these conditions (Fig. 5b, lane 3). However, addition to the translation reaction of nuclear extract from M059J cells (containing no DNA-PK activity) whose DNA had been damaged by γ -irradiation, together with purified holoenzyme, resulted in p53 activation (Fig. 5b, lane 5); nuclear extract from untreated M059J cells added with purified holoenzyme did not activate p53 DNA-binding (Fig. 5b, lane 4). These results indicate that DNA damage leads not only to DNA-PK activation, but also to the induction (independent of DNA-PK) of an additional factor(s), both of which are required for p53 activation. Results were similar when extracts from SCGR11 cells were used as the source of p53 (Fig. 5c). We conclude that activation of p53 DNA-binding requires both DNA-PK and an as-yet unidentified factor(s) present in nuclear extract from DNA-damaged cells.

Our data provide evidence that DNA-PK activates p53 in response to DNA damage so that p53 can bind to specific DNA sequences. As phosphorylation of p53 (at serines 15 and 37) by DNA-PK induced by ionizing radiation prevents MDM2 from inhibiting p53-dependent transactivation²¹, the action of DNA-PK on p53 seems to serve a dual purpose: it activates p53 DNA-binding while blocking its inactivation by MDM2. □

Methods

Cell culture and preparation of nuclear extracts. BALB/c 3T3 and SCGR11 cells were grown in DMEM supplemented with 10% fetal bovine serum (GIBCO-BRL). M059K and M059J cells were grown in DMEM-F12 supplemented with 10% fetal bovine serum (GIBCO-BRL). BALB/c or SCID mouse embryo fibroblasts were isolated from 13-day-old embryos²² and were cultured in DMEM supplemented with 10% fetal bovine serum. Cells were grown to 80% confluency and then mock-treated with phosphate-buffered saline (PBS) or treated with 200 nM hydrogen peroxide (BDH), 300 ng ml⁻¹ actinomycin D (CALBIOCHEM), or 10 μ g ml⁻¹ mitomycin C (SIGMA) in PBS for 18 h. For γ -irradiation, cells were grown to 80% confluency and then irradiated with a ¹³⁷Cs

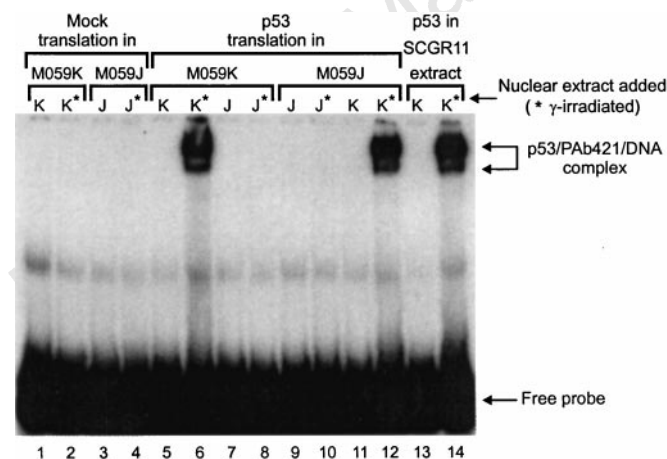
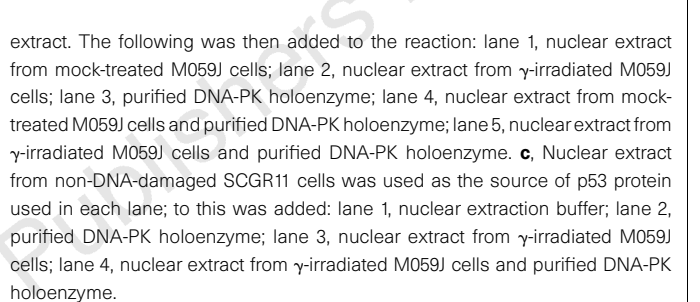


Figure 4 The effect of M059K and M059J nuclear extracts on DNA-binding activity of *in vitro*-synthesized p53 and SCGR11-cell p53. Cytoplasmic extracts from either untreated M059K or untreated M059J cells were used for either mock-translations (lanes 1–4) or translation of wild-type p53 (lanes 5–12). Following the *in vitro* translation reaction, an equal volume of nuclear extract immunodepleted of endogenous p53 was added as follows: lanes 1, 5 and 11, nuclear extract from non-DNA-damaged M059K; lanes 3, 7 and 9, nuclear extract from non-DNA-damaged M059J; lanes 2, 6 and 12, nuclear extract from γ -irradiated M059K; lanes 4, 8 and 10, nuclear extract from γ -irradiated M059J. Nuclear extracts from γ -irradiated cells are indicated by an asterisk. After addition of nuclear extracts to *in vitro* translation reactions, 10- μ l samples were tested for p53 sequence-specific DNA-binding activity by EMSA. The effect of nuclear extract from non-DNA-damaged (lane 13) or γ -irradiated (lane 14) M059K cells on DNA binding of p53 in SCGR11 cell extract was also examined by mixing equal volumes of the two nuclear extracts followed by EMSA.



extract. The following was then added to the reaction: lane 1, nuclear extract from mock-treated M059J cells; lane 2, nuclear extract from γ -irradiated M059J cells; lane 3, purified DNA-PK holoenzyme; lane 4, nuclear extract from mock-treated M059J cells and purified DNA-PK holoenzyme; lane 5, nuclear extract from γ -irradiated M059J cells and purified DNA-PK holoenzyme. **c.** Nuclear extract from non-DNA-damaged SCGR11 cells was used as the source of p53 protein used in each lane; to this was added: lane 1, nuclear extraction buffer; lane 2, purified DNA-PK holoenzyme; lane 3, nuclear extract from γ -irradiated M059J cells; lane 4, nuclear extract from γ -irradiated M059J cells and purified DNA-PK holoenzyme.

Polymerase chain reaction. The RT-PCR used to compare mRNA expression has been described²⁶. In brief, total cellular RNA was subjected to RT-PCR using random hexanucleotide primers (Pharmacia) and RTase (GIBCO-BRL) according to the manufacturers' protocol. The cDNAs from the RT-PCR step were then selectively amplified for p21 (CIP1) using the primers 5'-CGGTCCCGTGGACAGTGAGCAG-3' and 5'-GTCAGGCTGGTCTGCCTCG-3', which amplify a predicted 363-bp fragment. The GAPDH primers 5'-CGGAGTCAACGGATTGGTCGTAT-3' and 5'-AGCCTTCTCCATGGTGGTGAAGC-3' were used to amplify a predicted 306-bp GAPDH fragment which served as a PCR and gel-loading control. PCR products were analysed by

Cell-free *in vitro* translation. For cell-free *in vitro* translation, cytoplasmic extracts from M059K and M059J cells were prepared as reported²⁷, except that the extracts were not treated with micrococcal nuclease. Wild-type murine p53 mRNA was synthesized by *in vitro* transcription of a *Hind*III-linearized pSP6p53–Ala135 vector²⁸ using an Sp6 polymerase Megascript kit (Ambion). *In vitro* translation reactions were carried out in a final volume of 50 μ l containing the following components: 30 μ l cytoplasmic extract, 1 mM ATP, 0.5 mM GTP, 1 mg ml⁻¹ creatine phosphokinase (from ICN), 10 mM creatine phosphate, 40 mM haemin (from ICN), 80 mM KCl, 5 mM magnesium acetate, 1 mM DTT, 5 μ M of each amino acid (Promega), placental RNase inhibitor (RNA Guard, Pharmacia) and either water (for mock translations) or p53 mRNA (for p53 translations). Translation reactions were incubated at 37°C for 30 min, followed by the addition of an equal volume of nuclear extract from mock-treated or γ -irradiated M059K or M059J cells (see above) that had been immunodepleted of endogenous (probably mutant) p53 using the anti-p53 monoclonal antibody pAb421 preadsorbed onto inactivated *Staph. A* (IgSorb, The Enzyme Center). Reaction mixtures were then subjected to EMSA as described.

1. Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. Participation of p53 in the cellular response to DNA damage. *Cancer Res.* **51**, 6304–6311 (1991).
2. Tishler, R. B., Calderwood, S. K., Coleman, C. N. & Price, B. D. Increases in sequence-specific DNA binding by p53 following treatment with chemotherapeutic and DNA damaging agents. *Cancer Res.* **53**, 2212–2216 (1993).

3. Nelson, W. G. & Kastan, M. B. DNA strand breaks: The DNA template alterations that trigger p53-dependent DNA damage response pathways. *Mol. Cell. Biol.* **14**, 1815–1823 (1994).
4. Anderson, C. W. & Lees-Miller, S. P. The human DNA-activated protein kinase DNA-PK. *Crit. Rev. Eukaryot. Gene Expr.* **2**, 283–314 (1992).
5. Gottlieb, T. M. & Jackson, S. P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131–142 (1993).
6. Morozov, V. E., Falzon, M., Anderson, C. W. & Kuff, E. L. DNA-dependent protein kinase is activated by nicks and larger single-stranded gaps. *J. Biol. Chem.* **269**, 16684–16688 (1994).
7. Hendrickson, E. A. *et al.* A link between double-strand break-related repair and V(D)J recombination: The SCID mutation. *Proc. Natl Acad. Sci. USA* **88**, 4061–4065 (1991).
8. Blunt, T. *et al.* Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine SCID mutation. *Cell* **80**, 813–823 (1995).
9. Lees-Miller, S. P. *et al.* Absence of p350 subunit of DNA-activated protein kinase from a radiosensitive human cell line. *Science* **267**, 1183–1185 (1995).
10. Kern, S. E. *et al.* Oncogenic forms of p53 inhibit p53-regulated gene expression. *Science* **256**, 827–830 (1992).
11. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
12. El-Diery, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
13. Bogue, M. A., Zhu, C., Aguilar-Cordova, E., Donehower, L. A. & Both, D. B. p53 is required for both radiation-induced differentiation and rescue of V(D)J rearrangement in SCID mouse thymocytes. *Genes Dev.* **10**, 553–565 (1996).
14. Guidos, C. J. *et al.* V(D)J recombination activates a p53-dependent DNA damage checkpoint in SCID lymphocyte precursors. *Genes Dev.* **10**, 2038–2054 (1996).
15. Huang, L., Clarkin, K. C. & Wahl, G. M. p53-dependent cell cycle arrests are preserved in DNA-activated protein kinase-deficient mouse fibroblasts. *Cancer Res.* **56**, 2940–2944 (1996).
16. Nacht, M. *et al.* Mutations in the p53 and SCID genes cooperate in tumorigenesis. *Genes Dev.* **10**, 2055–2066 (1996).
17. Zhu, C., Bogue, M. A., Lim, D., Hasty, P. & Roth, D. B. Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. *Cell* **86**, 379–389 (1996).
18. Jhappan, C., Morse, H. C., Fleischmann, R. D., Gottesman, M. M. & Merlino, G. DNA-PK ζ : a T-cell tumour suppressor encoded at the mouse *scid* locus. *Nature Genet.* **17**, 483–486 (1997).
19. Hollstein, M. *et al.* Database of p53 gene somatic mutations in human tumors and cell lines. *Nucleic Acids Res.* **22**, 3551–3555 (1994).
20. Lees-Miller, S. P., Sakaguchi, K., Ullrich, S., Appella, E. & Anderson, C. W. Human DNA-activated protein kinase phosphorylates serines 15 and 37 in the amino-terminal transactivation domain of human p53. *Mol. Cell. Biol.* **12**, 5041–5049 (1992).
21. Shieh, S.-Y., Ikeda, M., Taya, Y. & Prives, C. DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell* **91**, 325–334 (1997).
22. Abbondanzo, S. J., Gadi, I. & Stewart, C. L. Derivation of embryonic stem cell lines. *Methods Enzymol.* **225**, 803–823 (1993).
23. McLure, K. G. & Lee, P. W. K. How p53 binds DNA as a tetramer. *EMBO J.* **17**, 3342–3350 (1998).
24. Chen, C. & Okayama, H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7**, 2745–2752 (1987).
25. Neumann, J. R., Morency, C. A. & Russian, K. O. A novel rapid assay for chloramphenicol acetyltransferase gene expression. *Biotechniques* **5**, 444–448 (1987).
26. Wong, H., Anderson, W. D., Cheng, T. & Riabowol, K. T. Monitoring mRNA expression by polymerase chain reaction: The “primer-dropping” method. *Anal. Biochem.* **223**, 251–258 (1994).
27. Skup, D. & Millward, S. Reovirus-induced modification of cap-dependent translation in infected L cells. *Proc. Natl Acad. Sci. USA* **77**, 152–156 (1980).
28. Milner, J., Medcalf, E. A. & Cook, A. C. Tumor suppressor p53: Analysis of wild-type and mutant p53 complexes. *Mol. Cell. Biol.* **11**, 12–19 (1991).

Acknowledgements. We thank P. Jeggo for SCGR11 SCID cells, J. Allalunis-Turner for M059J and M059K cells, J. Milner for the wild-type murine p53 plasmid pSP6p53-Ala, B. Vogelstein for pG₁₃CAT, J. Martinez for the pAb421 antibody, D. Chan for purifying DNA-PK, and E. Rattner and B. Carson for help with the preparation of mouse embryo fibroblasts. This work was supported by a grant (to P.W.K.L.) from the Alberta Cancer Board; K.G.M. was a recipient of the Alberta Heritage Foundation for Medical Research studentship.

Correspondence and requests for materials should be addressed to P.W.K.L. (e-mail: plee@acs.ualgary.ca).

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

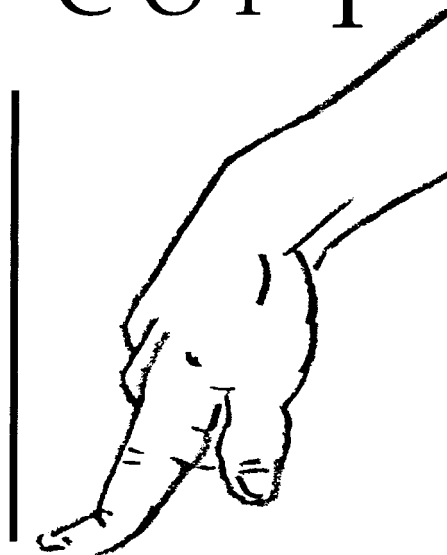
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

The software side of chemistry

Leading off with two chemical reaction stations, this roundup focuses mainly on software, including packages for chemometrics, the design of experiments, data acquisition and analysis, spectroscopy and remote Internet control.

Variomag

From H+P Labortechnik

Run up to 96 chemical reactions once

These heating and cooling reaction blocks allow fast heating to temperatures of over 200 °C and, if a cryostat is used, cooling to around -80 °C. The blocks should prove particularly useful in applications such as parallel organic synthesis, materials research and automation, as well as in biotechnology and combinatorial, analytical or metalloorganic chemistry. Furthermore, the reaction blocks are suited to integration with automated analytical equipment. A wear-free Variomag magnetic stirrer provides stirring without any moving parts. A stirring speed of 100–2,000 r.p.m. can be maintained with up to 96 samples stirred simultaneously. Bore holes in the reaction blocks are suitable for standard vessels, but the diameter and the depth of the bore holes can also be adjusted to non-standard vessels. A compact reflux cooler can be put onto the reaction block to prevent the evaporation of solvents. There is an inert gas cover to prevent the formation of condensate or ice. Moreover, the unit can be controlled manually or from a PC, and an RS-232 interface allows external control, documentation and recording of set values. Separation of reaction block and control unit makes these devices well suited for use in fume cupboards.

Reader Enquiry No. 100

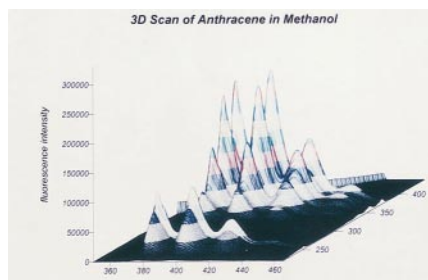
AccuCleave-96 workstation

From Irori

For high-throughput cleavage of up to 96 compounds synthesized on solid-phase supports

This cleavage station may be used interchangeably with either the AccuTag-100 combinatorial chemistry system or the automated AutoSort-10K system. Synthesis Manager software controls all of the systems and provides complete tracking of compound identification and archival locations. After cleavage, the cleaved compounds can be mapped into a variety of collection vial racks, which are designed around the standard microplate format. The station is said to have a highly reliable valveless design, eliminating the possibility of well-to-well cross contamination or leakage. All components are resistant to typical cleavage conditions, such as those using trifluoroacetic acid (TFA), dichloromethane and acetonitrile.

Reader Enquiry No. 101



A 'peak' at 3D fluorescence from Instruments S. A.

Three-dimensional fluorescence

From Instruments S.A.

A three-dimensional spectrofluorometer developed by Spex

With this instrument, a sample's complete spectral characterization of fluorescence spectrum can be produced. The optical design immediately disperses the excitation spectrum across one axis, the emission spectrum appears at right angles to this and the fluorescence intensity makes up the third dimension. As data are acquired instantly, millisecond-scale temporal data can be stored and analysed. DataMax software comes with the system and operates in a Windows environment.

Reader Enquiry No. 102

Remote control

TestPoint Internet/Intranet toolkit

From Keithley

Custom data-acquisition, instrument-control and analysis applications in Windows

This Internet toolkit can monitor and control tests or measurements from anywhere that an Internet connection can be established. The TestPoint web-server/e-mail object is said to be fast, simple and cost effective to operate — the user simply drags and drops it into an 'action list', and then chooses what to do. As the objects can be used within loops, conditionals and case statements, Internet connections and data transfers can be based on any set of conditions. The objects can also be used



Windows to the world: TestPoint from Keithley.

with timers, maths, data entry, graphics and other objects to provide timed connections and post only the needed data to the network. The toolkit helps to achieve complete remote control by allowing the user to view all or part of remote applications; it supports Intranet and Internet applications; views and captures graphs, pictures, audio and video; and accesses settings, buttons and other inputs. The system allows the use of 'server push' technology to automate remote updates and downloads, allowing users insert test data, images and graphs into web pages, retrieve e-mail updates from the test side, and have the test side send status updates to any list of e-mail addresses.

Reader Enquiry No. 103

Link for LIMS

From Contemporary Solutions

Version 3.5 of this software integrates a wide range of analytical instruments to LIMS

New features include optimized single-shot analyses, such as those typically used in quality-control and process-control environments. This includes viewing historical result on sample types, tighter integration of certain client/server LIMS and an extended range of configurable result manipulations. Other features include 'blank', 'drift' and 'recovery' corrections, as well as and rounding, limits and spreadsheet-type calculations. There is also manual and mixed manual/automatic data entry to LIMS, which can be included on a method-by-method basis. These are used increasingly to overcome the cumbersome nature of certain LIMS interfaces and to minimize the need for LIMS user licences. Multi-language capability allows immediate switching between operating languages.

Reader Enquiry No. 104

Instrumental assistance

Smart 1000

From Bruker AXS

A new 1K, CCD for the collection and solution of single-crystal diffraction data

This system uses a 1K, CCD chip with high sensitivity and deep, full-well capacity in X-ray diffraction applications. The CCD is bonded to a 90-mm diameter fibre-optics taper, with a 2.5:1 demagnification ratio — this has been combined with a custom phosphor screen. The measured gain at the molybdenum wavelength is a stated 28 electrons per X-ray photon. The new chip characteristics also allow cooling of the detector by a simple electronic Peltier device that

recirculates chilled water. Smart 1000 is typically configured with a 3-kW, sealed-tube, X-ray generator, a molybdenum X-ray tube and a platform goniometer, which allow the collection of data on single crystals of minerals, inorganic materials, metal cluster systems, organometallic complexes, organic compounds and natural products. At a typical 5-cm crystal-detector distance, a complete data set from 0° to 55° in 2 θ can be collected in a single detector setting for crystals with unit cell dimensions up to at least 50 Å. A copper source and custom phosphor screen (stated to be most efficient at the copper wavelength) can be used for the collection of organic and biological compounds. **Reader Enquiry No. 105**

VP Archive

From Shimadzuadzu

Data archiving software that is designed for chromatography laboratories

According to the manufacturer, VP Archive provides a fast, reliable and extremely secure method for archiving and retrieving data. Well suited for compliance purposes, it efficiently organizes and manages data, archiving data by user-defined, context-sensitive keys, such as study names or batch numbers. These keys provide fast access to data that may be days, months or even years old (thousands of files can be easily located with just a few mouse clicks). The software is fully automated and validated with a built-in audit trail, and it supports user-access control and log-on protocols.

Reader Enquiry No. 106

Experimental design

Design-Ease

From Stat-Ease

Windows software that is intended to aid in the design of experiments

This package should help users find critical factors affecting products and processes. Version 5 generates simple yet powerful, two-level factorial designs, as well as statistical and graphical outputs that should help reveal solutions to users. The software offers rotating three-dimensional plots and interactive two-dimensional graphics. It comes with free technical and statistical support, and an illustrated tutorial for easier learning.

Reader Enquiry No. 107

PLSplus/IQ Users

From Galactic

A spectroscopy chemometrics software package add-on to the Grams/32 package

This add-on program features multivariate analysis tools for building quantitative calibrations, as well as qualitative models for discriminant analysis. A 60-day trial period is offered, during which users can create

training sets of data, setup and run experiments for testing calibration models and predict unknown samples. Users can experiment with the program's preprocessing capabilities and easy-to-use 'report viewer', which assists in optimizing the calibration model and determining if there are outliers in training sets. Minimum system requirements for the trial version include Grams/32 version 5, a 486- or Pentium-compatible personal computer running Windows95 or NT 4.0, 20 MB of disk space, a math co-processor and a CD-ROM drive.

Reader Enquiry No. 108

Research and archiving

ChemSymphony Beans

From Cherwell Scientific

Java beans for chemists — a suite of tools for builders of Intranets with chemical content

This toolkit uses a common architecture and has been designed for object-oriented programming and data management. The first release consists of some 30 'beans', which perform many commonly used information processing tasks, such as drawing and editing molecular structures, generating search strings, translating file formats and presenting the results of complex searches. Beans for the visualization, interaction with molecular structures and connectivity with other Intranet resources are also provided. ChemSymphony Beans adheres to JavaSoft's JDK 1.1 standards can be customized and linked to beans from third parties. Other features include graphics, Java code, accessible properties, two- and three-dimensional renderers for viewing single and multiple structures, control/interface beans, and flat-file (text) and chemical-property displays.

Reader Enquiry No. 109

SciFinder Research

From Chemical Abstract Services

A service that aids in research, the browsing of scholarly journals and keeping up to date on the most recent scientific developments

Version 3.0 includes 'Refine', a component that allows the user to pinpoint a structure or reference-answer set by several different criteria. Users can explore by a structure and refine it by another structure or by commercial availability. Refining by reference is also possible, as is exploration by research topic, author name or through the use of 'keep-me-posted' alerts. Users can also refine an answer set by author name, research topic, document type, publication year and language. There are features that allow the user to save an answer set and open the set later, without re-executing the exploration. SciFinder Scholar is a special version of this desktop research tool. According to the manufacturer, students can begin research in minutes with little or no training.

SciFinder Scholar is available for campus-wide access, but has limited access times.

Reader Enquiry No. 110

Cerius2 version 3.6

From Molecular Simulations

Recently enhanced library diversity analysis tools for use in combinatorial chemistry

Two new modules, C2*LibCompare for library comparison and C2*LibSelect for combinatorial selection of reagents, now complement this suite of molecular diversity tools. Additional enhancements include access to the company's ISIS Fingerprints and Daylight Fingerprints. There is support for multidimensional scaling, clustering and diversity-selection schemes for fingerprint descriptors, Kier and Hall electro-topological state descriptors, factor analysis and the integration of Catalyst/SHAPE descriptors for structure base focusing.

Reader Enquiry No. 111

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

PIERCE

SuperSignal® BLAZE™
Chemiluminescent Substrate*

The Ultimate Sensitivity for Western Blotting

Features/Benefits:

- Ultrahigh femtogram-level (10⁻¹⁵) sensitivity
- Excellent signal-to-noise ratios
- Emits a strong signal for up to eight hours
- Recommended primary antibody dilution is 1:5,000-1:100,000; recommended secondary antibody dilution is 1:100,000-1:500,000
- Eight-hour working solution stability
- Allows for stable hard-copy results on film, rapid processing and high resolution

Tel: 800-874-3723 or 815-968-0747 • Fax: 815-968-7316
CS@piercenet.com • Internet: www.piercenet.com

* Patent Pending

READER ENQUIRY NO.

Silver Stain Kit II for Electrophoresis

**Rapid, sensitive detection
of proteins & nucleic acids**



Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO.